

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:09 ; Search time 19.0439 Seconds

(without alignments)  
1661.397 Million cell updates/sec

Title: US-10-085-167-2

Perfect score: 1726  
Sequence: 1 MLEPULGLGPACWALGPT.....LVTPDLAPAPPGIGASELL 329

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR.76:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	213	12.3	219	2 T14782	hypothetical prote
2	196	11.4	244	2 JC4708	gelatin-binding 28
3	184	10.7	215	2 B48150	hibernation-relate
4	182	10.5	253	1 C1H0B	complement subcomp
5	176.5	10.2	253	2 S49158	complement protein
6	175	10.1	744	1 S23298	collagen alpha 1(V
7	173	10.0	744	1 A34246	collagen alpha 1(V
8	172	10.0	245	1 C1H0C	complement subcomp
9	170	9.8	744	2 S15435	collagen alpha 1(V
10	169.5	9.8	674	2 S23297	collagen alpha 1(X
11	168.5	9.8	680	2 S31216	collagen alpha 1(X
12	168	9.7	245	2 S19018	complement subcomp
13	166.5	9.6	680	1 CGH0D	collagen alpha 1(X
14	165.5	9.6	743	1 S23779	collagen alpha 1(V
15	164.5	9.5	253	2 I49560	complement C1q B C
16	163.5	9.5	674	2 S13301	collagen alpha 1(X
17	163	9.4	224	2 A60032	ceredellin-like g1
18	160.5	9.3	246	2 S29338	complement subcomp
19	158	9.2	193	2 A37873	ceredellin precurs
20	151.5	8.8	245	1 C1H0A	collagen alpha 1(X
21	150	8.7	196	2 A48150	hibernation-relate
22	147.5	8.5	215	2 C48150	hibernation-relate
23	147	8.5	1228	2 A57384	multimerin, endoth
24	145.5	8.4	423	2 A57397	collagen precursor
25	139.5	8.1	635	2 A57131	collagen alpha 2(V
26	135.5	7.9	170	2 B57131	collagen alpha 2(V
27	131.5	7.6	992	2 T08772	hypothetical prote
28	98	5.7	3938	2 T42761	Bassoon protein -
29	97.5	5.6	476	2 T29083	3-isopropylmalate

30	95	5.5	495	1 G64878	aldehyde dehydroge
31	94.5	5.5	619	2 A87499	hypothetical prote
32	94	5.4	495	2 E90863	aldehyde dehydroge
33	94	5.4	495	2 D85755	aldehyde dehydroge
34	94	5.4	810	2 B84185	cytochrome-like pr
35	94	5.4	913	2 AB2587	translation initia
36	94	5.4	913	2 B97369	translation initia
37	94	5.4	1045	2 A39199	endoglucanase B (E
38	94	5.4	7463	2 T36248	CDA peptide synthe
39	93.5	5.4	686	1 VGBENA	glycoprotein H pre
40	93	5.4	462	2 B75306	probable cyclotrope
41	93	5.4	481	2 B75480	hypothetical prote
42	93	5.4	1239	1 Q08B10	BOLFI protein - hu
43	92.5	5.4	420	2 E75130	hypothetical prote
44	92.5	5.4	1334	2 T50568	probable multi-dom
45	91.5	5.3	13055	2 T16580	hypothetical prote

#### ALIGNMENTS

RESULT 1  
T14782  
hypothetical protein DKFZ586B0621.1 - human (fragment)  
C/Species: Homo sapiens (man)  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 02-Sep-2000  
C/Accession: T14782  
R/Ottensmeyer, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
Submitted to the Protein Sequence Database, August 1999  
A/Reference number: Z18184  
A/Accession: T14782  
A/Status: Preliminary  
A/Molecule type: mRNA  
A/Residues: 1-219 <OTT>  
A/Cross-references: EMBL:AL110261  
A/Experimental source: adult uterus; clone DKFZ586B0621  
C/Genetics:  
A/Note: DKFZ586B0621.1  
C/Superfamily: complement C1q carboxyl-terminal homology

Query Match 12.3%; Score 213; DB 2; Length 219;  
Best Local Similarity 34.9%; Pred. No. 4.3e-10;  
Matches 66; Conservative 18; Mismatches 83; Indels 22; Gaps 6;

QY	138	GAPHYALGAGATFSG---YLVIADAD---ADAPARGPAPP-----EPSSAFSARTR 185	
DB	33	GAP---GAPGEGEGRGRLPGPRGDPGPRGAGAGPFGAGGCVPPRASFSAKRSE 88	
QY	186	SLVGS DAGPGRHQPLAFTEFVNIGDPAAGVRCRLPGAYFFSFTLGKLPKRTLSV 245	
DB	89	SRV-----PPSPAPLPFPRVLVNEQHYDAVTKFTCCQPGVYFA-VHATVYRASLGF 142	
QY	246	KLMKNDVQAMTYDDGASRRRMOQSVMALRRDPAWLSHDHGGAGVSNHGKITYT 305	
DB	143	DLVKNGESIASFPQFGWPKPASLSGAMVRLBPDQWVGVDYIGIVASIKTIDST 202	
QY	306	FSGFVLYPPD 314	
DB	203	FSGFVLYSD 211	

RESULT 2  
JC4708  
gelatin-binding 28k protein precursor - human  
N/Alternate names: adipose specific collagen-like factor  
C/Species: Homo sapiens (man)  
C/Date: 10-May-1996 #sequence\_revision 19-Jul-1996 #text\_change 20-Sep-1999  
C/Accession: JC4708; JC4944  
R/Maeda, K.; Okubo, K.; Shimomura, I.; Funahashi, T.; Matsuzawa, Y.; Matsubara, K.  
Biochem. Biophys. Res. Commun. 221, 286-289, 1996  
A/Title: cDNA cloning and expression of a novel adipose specific collagen-like factor,  
A/Reference number: JC4708; MIMD:96324171; PMID:8619847  
A/Accession: JC4708

A: Molecule type: mRNA  
A: Residues: 1-244 <MAE>  
A: Cross-references: DDBJ:ID45371; NID: g871886; PIDD: BAA08227.1; PID: g871887  
A: Experimental source: adipose tissue  
J: Nakano, Y.; Tobo, T.; Choi-Makura, N. H.; Mazda, T.; Tomita, M.  
J. Biochem. 120, 803-812, 1996  
A: Title: Isolation and characterization of GBP28, a novel gelatin-binding protein purified from sheep placenta  
A: Reference number: J49494; MUID: 97103474; PMID: 8947845  
A: Accession: J49494  
A: Molecule type: protein  
A: Residues: 19-38; 93-100; 101-112; 135-149; 173-178 <NAK>  
C: Comment: This protein is an endogenous factor that binds with a collagen-like domain.  
C: Genes: C1q  
A: Gene: apm1  
C: Superfamily: unassigned collagens; complement C1q carboxyl-terminal homology  
C: Keywords: adipose tissue; glycoprotein; hydroxyproline  
F: 1-18/Domain: signal sequence #status predicted <SIG>  
F: 19-244/Product: gelatin-binding 28kDa protein #status experimental <MAT>  
F: 142-107/Region: collagen-like  
F: 114-241/Domain: complement C1q carboxyl-terminal homology <C1Q>  
F: 95/Modified site: 4-hydroxyproline (Pro) #status experimental  
F: 230/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 11.4%; Score 196; DB 2; Length 244;  
Best Local Similarity 29.7%; Pred. No. 1.2e-08;  
Matches 54; Conservative 31; Mismatches 61; Indels 36; Gaps 8;

QY 7 GLUGP-----AACWALGP-----TPPGSGSELSAFSAARTPLEGTSMAV 48  
DB 72 GLIGPKGIDIGETGVGAGEGPRGPGIQRKKGPEGAYVRSVGLETYVT-IPNMP 130  
QY 49 TPDKRVYVIGDEPVATQGFRCRVPGARVFSEFTGAKPH-----KSLSVMLVRNDEVQA 103  
DB 131 RFTKIFVYQNNHYDSTKFRHCNIPGLYFAY-----HITVYMKDVKVSIFK-KDKAML 183  
QY 104 LAPDEQRPRGARBAASGASMLQDYGDTVRLRGAPH---VALGAPGATFGSYLVYAD 159  
DB 184 FTYDDYQENNVDQ-ASSGSVLHLHEVDQVWLQVIGEGRNGLYADNDNDSTFTFLYHD 242

QY 160 AD 161  
DB 243 TN 244

RESULT 3  
B48150  
hibernation-related protein HP-25 precursor - Siberian chipmunk  
C: Species: Eutamias sibiricus (Siberian chipmunk)  
C: Date: 16-Feb-1994 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
C: Accession: B48150; B41752  
R: Takamatsu, N.; Ohba, K.; Kondo, J.; Kondo, N.; Shiba, T.  
Mol. Cell. Biol. 13, 1516-1521, 1993  
A: Title: Hibernation-associated gene regulation of plasma proteins with a collagen-like  
A: Reference number: A48150; MUID: 93180798; PMID: 8441393  
A: Accession: B48150  
A: Molecule type: mRNA; protein  
A: Residues: 1-215 <TAKE>  
A: Cross-references: GB:ID12975; NID: g287469; PID: BAA02352.1; PID: g287470  
A: Note: the source is designated as Tamias asiaticus in Genbank entry TMSHP25, release 1  
A: Note: sequence extracted from NCBI backbone (NCBIN:125946, NCBIPI:125947)  
R: Kondo, N.; Kondo, J.  
J. Biol. Chem. 267, 473-478, 1992  
A: Title: Identification of novel blood proteins specific for mammalian hibernation.  
A: Reference number: A41752; MUID: 92112696; PMID: 1730610  
A: Accession: B41752  
A: Status: preliminary  
A: Molecule type: protein  
A: Residues: 29-62; 94-130; 172-183; 187-192; 201-215 <KON>  
C: Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal homology  
C: Keywords: glycoprotein; hibernation; plasma  
F: 1-28/Domain: signal sequence #status predicted <SIG>  
F: 29-215/Product: hibernation-related protein HP-25 #status experimental <MAT>  
F: 91-214/Domain: complement C1q carboxyl-terminal homology <C1Q>

F;167/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 10.7%; Score 184; DB 2; Length 215;  
Best Local Similarity 30.4%; Pred. No. 9,6e-08;  
Matches 62; Conservative 22; Mismatches 52; Indels 68; Gaps 11;

Oy 138 GAPHYALGAGATSGVLLYADADAPAGPAPAP-----EPRS 177  
Db 49 GIFGFPG-GAPQAL-----GPEPPFVPGLPGHQGPDPVEKSSRPKS 90

Oy 178 AFSAASTRSLVSGDAGPGRHOPLAFPTFEVNTIGDFDAAGVRCPLPGAYPEFTLKG 237  
Db 91 AFAYKL-----SERPEPF-FQPIVFKEALYNQGHFMATGERSCVLPGVYNGFDI-R 142

Oy 238 LPKRLLTSVKLMKNRDEVOAMITDDGASRRRMOSQ-----SYMLALRGDAWLLSH 289  
Db 143 LFQSSVXIRILM--RDGIQV-----REKBAQANDSKHMGSVIALGDKGVWLDSK 192

Oy 290 DHDGYGAVSNHG-KYIFSGFLVY 312  
Db 193 LK---GTSEKKTTHIVFGILTY 213

RESULT 4  
C1IHOB  
Complement subcomponent C1q chain B precursor (validated) - human  
N/Alternate names: complement subcomponent C1q beta chain  
C/Species: Homo sapiens (man)  
C/Date: 22-May-1991 #sequence revision 31-May-1996 #text change 08-Dec-2000  
C/Acession: B23442; A23442; B90304; A90301; B90315; A03206  
R/Reid, K.B.M.  
Biochem. J. 231, 729-735, 1985  
A>Title: Molecular cloning and characterization of the complementary DNA and gene coding  
A/Reference number: A23442; MUID:86076906; PMID:3000358  
A/Acession: B23442  
A/Molecule type: DNA  
A/Residues: 'HS',1-32 <RE1>  
A>Note: The authors translated the codon AGT for the second position as Arg; they were  
A/Acession: A23442  
A/Molecule type: mRNA  
A/Residues: 28-253 <RE1>  
A/Cross-references: EMBL:X03084  
A>Note: the authors translated the codon ACA for residue 46 as Ile  
R/Reid, K.B.M.  
Biochem. J. 179, 367-371, 1979  
A>Title: Complete amino acid sequences of the three collagen-like regions present in su  
A/Reference number: A90304; MUID:80020137; PMID:486087  
A/Acession: B90304  
A/Molecule type: protein  
A/Residues: 'E',29-84,'D',86-99,'P',101-135 <RES>  
R/Reid, K.B.M.; Thompson, E.O.P.  
Biochem. J. 177, 863-868, 1978  
A>Title: Amino acid sequence of the N-terminal 108 amino acid residues of the B chain o  
A/Reference number: A90301; MUID:79041552; PMID:708376  
A/Acession: A90301  
A/Molecule type: protein  
A/Residues: 28-99,'P',101-195 <RE3>  
R/Reid, K.B.M.; Gagnon, J.; Frampton, J.  
Biochem. J. 203, 559-569, 1982  
A>Title: Completion of the amino acid sequences of the A and B chains of subcomponent C  
A/Reference number: A90315; MUID:82283890; PMID:6981411  
A/Acession: B90315  
A/Molecule type: protein  
A/Residues: 136-253 <RE4>  
A>Note: 176-Glx may also be present  
C/Comment: The first component of complement is a calcium-dependent complex of the thre  
vation of C1r (enzyme), C1s (proenzyme), and the other eight components of complement  
C/Comment: The C1q subcomponent is composed of nine subunits, six of which are disulfid  
(see PIR:C1HUQC) chain. Equimolar amounts of the A, B, and C chains are found after red  
C/genetics:  
A/Gene: GDB:C1OB  
A/Cross-references: GDB:119043; OMIM:120570  
A/Map position: lp36.3-lp34.1

C/Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal hom  
 C/Keywords: complement pathway; glycoprotein; heterodimer; hydroxylysine; hydroxyproline  
 F:1-27/Domain: signal sequence #status predicted <SIG>  
 F:28-253/Product: complement subcomponent C1q chain B #status experimental <MAT>  
 F:3-116/Domain: collagenous, triple helix <COL>  
 F:123-249/Domain: complement C1q carboxyl-terminal homology <C1Q>  
 F:28/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experim  
 F:31/Dsulfide bonds: interchain (to chain A-26) #status experimental  
 F:35,38,41,53,56,65,83,86,101,104,107/Modified site: 4-hydroxyproline (Pro) #status exp  
 F:59,62,77,92,98,110/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F:59,62,98,110/Binding site: carbohydrate (Lys) (covalent) #status experimental

Query Match 10.5%; Score 182; DB 1; Length 253;  
 Best Local Similarity 29.1%; Pred. No. 1,7e-07;

Matches 48; Conservative 32; Mismatches 61; Indels 24; Gaps 5;

QY 165 PARGPAPAPP-----RSAPSAFTSLVSGDAGCPRRQPLAFTEFNIG 211  
 DB 97 PKGPGAPGAPGKSGSGDKATQKIAFSATRLIN-----PLRRQTRFPHVITNM 150  
 QY 212 GDFDAAGVFRCLPGAYFSPFTLGKLPKRLTSVKLMKNDEVQAMI-YDDGASRRREM 270  
 DB 151 NNNEPSSGKFTCKVPGHYFTYHASS--RGNLCVNLRRGRERQKVTFCDVAYNPTQVT 208  
 QY 271 SQSVMLALRRGDVAVWLSHHDGAYGAYSNHGKXITTSGLVLPDL 315  
 DB 209 TGGWVLKLEGGENVFLQATDKN--SLTGMGANSIFSGFLFPDM 251

## RESULT 5

S49158  
 Complement protein C1q beta chain precursor - rat

C/Species: Rattus norvegicus (Norway rat)  
 C/Date: 16-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 20-Aug-1999

C/Accession: S49158  
 R/Schwaebel, W.; Petry, F.; Loos, M.

submitted to the EMBL Data Library, March 1993

A/Description: cDNA cloning and expression of the mRNA encoding the B-chain of rat C1q.

A/Reference number: S49158

A/Accession: S49158

A/Status: Preliminary

A/Molecule type: mRNA

A/Residues: 1253 <SCH>

A/Cross-references: EMBL:X1127; NID:G510191; PIDN:CAA50440.1; PID:G510192

C/Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal hom  
 F:121-249/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 10.2%; Score 176.5; DB 2; Length 253;  
 Best Local Similarity 28.2%; Pred. No. 4.8e-07;

Matches 50; Conservative 40; Mismatches 66; Indels 21; Gaps 8;

QY 2 LPULLGLGPAACWALGP--TPGP-----GSSELRS---AFSARTTPEGTSMA 47  
 DB 80 IPGIPKGVKPKG--PVGPKAPGPPGPRGKSGSGDKATQKIAFSATRLINVSALPNQA 137  
 QY 48 VTFDKYVYVNGDFVATGQFCRCVPGAYFSPFTAGKAPKSLVWLVRNRDEVQ-ALAF 106  
 DB 138 IREKIVITVNDVYEPSSGKFTCKVPGHYFTYHASSRGNLCVNLRRGRERDMQVLF 197  
 QY 107 DEGRRCGARRAASQSMQLQDLYGDTYWLRLHGAPHYALGAPAT--FSGYLVAADD 161  
 DB 198 CDYAAQ-NTFQVTTGGVVLKLEQEVVHLQATD-KNSLIGVEGANSIFTGFLFPDM 252

## RESULT 6

S23298  
 collagen alpha 1(VIII) chain - chicken

C/Species: Gallus gallus (chicken)

C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C/Accession: S23298  
 R/Ninomiya, Y.; Castagnola, P.; Gerecke, D.; Gordon, M.K.; Jacenko, O.; LuValle, P.; McC  
 maguchi, N.; Olsen, B.R.  
 in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp.79-114, Academic Pre

A/Title: The molecular biology of collagens with short triple-helical domains.  
 A/Reference number: S22243  
 A/Accession: S23298  
 A/Status: Preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-744 <NIN>  
 C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
 F:617-743/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 10.1%; Score 175; DB 1; Length 744;  
 Best Local Similarity 25.3%; Pred. No. 2.4e-06;

Matches 56; Conservative 28; Mismatches 77; Indels 60; Gaps 8;

QY 136 LHGAPHYALGAPGATFSGYLVAADADADA-RRGAPAPPEPR- 177  
 DB 540 LHGPP-----GKPGA-----LGPQGPGLPGPPGPPGPPAPVMPPTAPQGEYLPDMG 589  
 QY 178 -----AFSAARTSLVSGDAGP-----PRHQPLAFTEFNIG 212  
 DB 590 LGIDGVKTPHAYAAKK-----GKNGGPAVEMPAFTALTAPEPPVGAFTKFDRLVNGRQ 644  
 QY 213 DFDAAAGVFRCLPGAYFSPFTLGKLPKRLTSVKLMKNDEVQAMIYDDGASRRREM 272  
 DB 645 NNNEPSSGKFTCKVPGHYFTYHASS--RGNLCVNLRRGRERQKVTFCDVAYNPTQVT 208  
 QY 273 SQSVMLALRRGDVAVWLSHHDGAYGAYSNHGKXITTSGLVLPDL 315  
 DB 703 SAVLLLRPGDRVFLQNPSEQAAGLVAGYVHSSFSGLLYP 743

## RESULT 7

A34246

collagen alpha 1(VIII) chain precursor - rabbit

C/Species: Oryctolagus cuniculus (domestic rabbit)  
 C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C/Accession: A34246  
 R/Yamaguchi, N.; Benya, P.D.; van der Rest, M.; Ninomiya, Y.

J. Biol. Chem. 264, 16022-16029, 1989  
 A/Title: The cloning and sequencing of alpha1(VIII) collagen cDNAs demonstrate that tye  
 remains similar to those of type X collagen.

A/Reference number: A34246; MUID:89380159; PMID:2476437

A/Accession: A34246

A/Molecule type: mRNA

A/Residues: 1-744 <YAM>

A/Cross-references: GB:J05042; NID:G164895; PIDN:AAA11204.1; PID:G164896

C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
 F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-744/Product: collagen alpha 1(VIII) chain #status predicted <MAT>

F:118-571/Region: amino-terminal nonhelical

F:572-744/Region: interrupted helical

F:617-743/Domain: complement C1q carboxyl-terminal nonhelical

Query Match 10.0%; Score 173; DB 1; Length 744;  
 Best Local Similarity 25.3%; Pred. No. 3.4e-06;

Matches 56; Conservative 27; Mismatches 78; Indels 60; Gaps 8;

QY 136 LHGAPHYALGAPGATFSGYLVAADADADA-RRGAPAPPEPR- 177  
 DB 540 LHGPP-----GKPGA-----LGPQGPGLPGPPGPPGPPAPVMPPTAPQGEYLPDMG 589  
 QY 178 -----AFSAARTSLVSGDAGP-----PRHQPLAFTEFNIG 212  
 DB 590 LGIDGVKTPHAYAAKK-----GKNGGPAVEMPAFTALTAPEPPVGAFTKFDRLVNGRQ 644  
 QY 213 DFDAAAGVFRCLPGAYFSPFTLGKLPKRLTSVKLMKNDEVQAMIYDDGASRRREM 272  
 DB 645 NNNEPSSGKFTCKVPGHYFTYHASS--RGNLCVNLRRGRERQKVTFCDVAYNPTQVT 208  
 QY 273 SQSVMLALRRGDVAVWLSHHDGAYGAYSNHGKXITTSGLVLPDL 315  
 DB 703 SAVLLLRPGDRVFLQNPSEQAAGLVAGYVHSSFSGLLYP 743

## RESULT 8

complement subcomponent C1q chain C precursor - human  
 N:Alternate names: complement subcomponent C1q gamma chain  
 C:Species: Homo sapiens (man)  
 C>Date: 22-May-1991 #sequence\_revision 31-May-1996 #text\_change 22-May-1998  
 C/Accession: S14351; A03207  
 R:Seller, G.C.; Blake, D.J.; Reid, K.B.M.  
 Biochem. J. 274, 481-490, 1991  
 A>Title: Characterization and organization of the genes encoding the A-, B- and C-chains  
 A/Reference number: S14350; MUID:91174759; PMID:1706597  
 A/Accession: S14351  
 A>Status: not compared with conceptual translation  
 A:Molecule type: DNA  
 A/Residues: 1-245 <SEU>  
 R:Reid, K.B.M.  
 Biochem. J. 179, 367-371, 1979  
 A>Title: Complete amino acid sequences of the three collagen-like regions present in sub  
 A/Reference number: A90304; MUID:80020137; PMID:486087  
 A/Accession: A03207  
 A:Molecule type: protein  
 A/Residues: 29-56, 'P', 58-65, 'K', 67-71, 'P', 73-83, 'K', 85-86, 'D', 88-89, 'N', 91-122 <REI>  
 C/Comment: The first component of complement is a calcium-dependent complex of the three  
 C:Comment: of C1r (enzyme), C1s (proenzyme), and the other eight components of complement.  
 C:Comment: The C1q subcomponent is composed of nine subunits, six of which are disulfide  
 C:Comment: dimers of the C chain. Equimolar amounts of the A, B, and C chains are found after red  
 C/Genetics:  
 A:Gene: GDB: C1QG  
 A/Cross-references: GDB:128132; OMIM:120575  
 A:Map position: 1p36.3-1p34.1  
 A:Introns: 60/3  
 C:Superfamily: complement subcomponent C1q chain A: complement C1q carboxyl-terminal hom  
 C:Keywords: complement pathway; glycoprotein; homodimer; hydroxylysine; hydroxyproline;  
 F:1-28/Domain: signal sequence #status predicted <SIG>  
 F:29-245/Product: complement subcomponent C1q chain B #status predicted <WAT>  
 F:31-114/Domain: collagenous, triple helix <COL>  
 F:121-244/Domain: complement C1q carboxyl-terminal homology <C1Q>  
 F:32/Disulfide bonds: interchain #status experimental  
 F:36,39,42,45,54,63,81,93,96,99,105/Modified site: 4-hydroxyproline (Pro) #status experi  
 F:57,72,75/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F:75/Binding site: carboxylate (Lys) (covalent) #status experimental

Query Match 10.0% Score 172; DB 1; Length 245;

Best Local Similarity 23.8%; Pred. No. 1.1e-06;

Matches 67; Conservative 33; Mismatches 84; Indels 98; Gaps 13;

QY 86 PHKSLVNLVNRDEVQALAFDEQRPGARRAASQASAMLDYDPTWLRHGAPHALG 145  
 DB 9 PHLGAKLL-----LILLALRGQANTG-----YG-----IPGMGLP-G 43  
 QY 146 APGATFSGY--LVVADADADA-----RGPAPR----- 173  
 DB 44 AFGK--DGYDGLPGKGGPGIPALGIRPGKQKGPGLPGHKGKGMGPGMGVGP 101  
 QY 174 -----EP-----RSAPFARTSLVSDAGPGRHOP-----LAFDTEFVNTGG 212  
 DB 102 MGIPGEBEGRYKQKFGSVFTVTR-----QTHQPPANSLIRPAVLTINQG 149  
 QY 213 DFDAAGVFRCLPGAYFFSFTLGLPKRTLSVKLMKNRDEVQAMLYDQASRRREMOSQ 272  
 DB 150 DYDSTGTGKTKVPGVLYVYVNAS-----HTANLCVLLVRSGVKVVTFCGHTSKTNQVNSG 205  
 QY 273 SYMLALRRGDVAVLSSHHDGCGAYSNHGKTYTFSGFLVYP 314  
 DB 206 GVLRLQVGEVWLAIVNDY--YDMVGIQGSDFVSGFLFLFP 245

## RESULT 9

S15435  
 collagen alpha 1(VIII) chain precursor - human  
 C:Species: Homo sapiens (man)  
 C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 13-Aug-1999

C/Accession: S15435

R:Muragaki, Y.; Mattei, M.G.; Yamaguchi, N.; Olsen, B.R.; Ninomiya, Y.  
 Eur. J. Biochem. 197, 615-622, 1991

A>Title: The complete primary structure of the human alpha-1(VIII) chain and assignment  
 A/Reference number: S15435; MUID:91231001; PMID:2029894

A/Accession: S15435

A>Status: preliminary

A:Molecule type: mRNA

A/Residues: 1-744 <MR>

A/Cross-references: EMBL:X57527; NID:G30081; PID:CAA40748.1; PID:G30082

C/Genetics:

A:Gene: GDB: COL8A1

A/Cross-references: GDB:128104; OMIM:120251

A:Map position: 3q11.1-3q13.2

C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-744/Product: collagen alpha 1(VIII) chain #status predicted <WAT>

F:21-117/Region: amino-terminal nonhelical

F:118-571/Region: interrupted helical

F:572-744/Region: carboxyl-terminal nonhelical

F:617-743/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 9.8% Score 170; DB 2; Length 744;

Best Local Similarity 25.9%; Pred. No. 6e-06;

Matches 56; Conservative 27; Mismatches 83; Indels 50; Gaps 9;

QY 136 LHGAPHYALGAPATFSGYLVVADADADA--RGPAPRPPRS----- 177  
 DB 540 LHGP-----GKPA-----LGPQGQGLPGPPGPPPPPAVMPPTPPGGLYEDMG 589  
 QY 178 -AFSAARTSLVSDAGP--GPRHPLAPTE---FVNIG-----DFDA 217  
 DB 590 LGIDGVPPPHATGAKKGNKGNPVEMPAFYAEVLTAFPPPGGVKFNKLYNGRONVPG 649  
 QY 218 AGVFRCLPGAYFFSFTLGLPKRTLSVKLMKNRDEVQAMLYDQASRRREMOSQVLA 277  
 DB 650 TGIFTCVPPVYFAVHV--HCKGNVWALFKNNPVM--YTYBYKKGFIDQASGSAYLL 707  
 QY 278 LRRGDVAVLSSHHDGCGAYSNHGKTYTFSGFLVYP 313  
 DB 708 LRRGDVAVLSSHHDGCGAYSNHGKTYTFSGFLVYP 313

## RESULT 10

S23297  
 collagen alpha 1(X) chain precursor - chicken

N:Alternate names: type X collagen  
 C:Species: Gallus gallus (chicken)

C>Date: 07-Oct-1994 #sequence\_revision 10-Nov-1995 #text\_change 13-Aug-1999

C/Accession: S23297; A31896; S65594; S77711; I50218

R:Ninomiya, Y.; Castagnola, P.; Gerecke, D.; Gordon, M.K.; Vacenko, O.; LuValle, P.; Mc  
 maguchi, N.; Olsen, B.R.

A>Title: The molecular biology of collagens with short triple-helical domains.  
 A/Reference number: S23297

A/Accession: S23297

A>Status: preliminary

A:Molecule type: mRNA

A/Residues: 1-674 <NIN>

R:LuValle, P.; Ninomiya, Y.; Rosenblum, N.D.; Olsen, B.R.

J. Biol. Chem. 263, 18378-18385, 1988

A>Title: The type X collagen gene. Intron sequences split the 5'-untranslated region an

A/Reference number: A31896; MUID:89054019; PMID:2461368

A/Accession: A31896

A:Molecule type: mRNA

A/Residues: 1-75 <LVV>

R:Ninomiya, Y.; Gordon, M.; van der Rest, M.; Schmid, T.; Linsemayer, T.; Olsen, B.R.

J. Biol. Chem. 261, 5041-5050, 1986

A>Title: The developmentally regulated type X collagen gene contains a long open readin

A/Reference number: I50218; MUID:86168227; PMID:3082876

A/Accession: S65594

A:Molecule type: DNA

A/Residues: 'T', '9', 'D', '11-12', 'EDOMKLYILFTW', '30-31', 'TCKSGRAFTYIIONVADLVSSH', '48-89', 'L'



```

629. 'POAVLSISMRITKCGSSCOLONPMVSIPLMFTILSOVSYLKSNIP.LTMS' <NIN1>
A:Cross-references: EMBL:M13496; NID:G211699; PIDN:AAA48736.1; PID:G211700
A:Accession: S177711
A:Molecule type: Protein
A:Residues: 104-112, 'X', 114-117, 453-466 <NIN2>
C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology
C:Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer; hydroxyproline;
F:1-18/Domain: signal sequence #status predicted <SIG>
F:547-673/Domain: complement C1q carboxyl-terminal homology <C1Q>
F:453,456/Modified site: hydroxyproline (Pro) #status experimental
F:611/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match
Best Local Similarity 25.3%; Score 169.5; DB 2; Length 674;
Matches 58; Conservative 21; Mismatches 75; Indels 75; Gaps 7;

QY 145 GAGCATFSGYLTVYADADADADA-----RGPAPPEPR----- 176
Db 454 GAGC-----AKGEAGAGLPGPAGIATKGLRGPMPGPPGPKGNSGEPGLPDPG 504
QY 177 -----SAPSAARTSLVQ-----SDAGPPRRHQL 201
Db 505 PPGPGQSTIEGYVKGSRRLSGMSFMKAGANALGMPVSATVLISKAYPGAT-VPI 563
QY 202 AFDFEYVNIIGDPPAAAGVFCRLPGATFFSFTLGKLPKRLSYLKMKNDEVQAMYYDD 261
Db 564 KFDILYLRQOHYDPRGTIFTCRIPGLYFSYHV-HAKGTWVALYKNGSPVW-YTYDE 621
QY 262 GASRRMOSQSVMLALRGAVALSHDHGCGANSNGKYITFGSGL 310
Db 622 YOKGYLDOAGSNAVITDLMENDQWMLQLPNSNGLYSEYVHSSFGSL 670

RESULT 11
S31216
collagen alpha 1(X) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 13-Aug-1999
C:Accession: S31216; S28807; S22215; S30127; 148299; S26397; S31830
R:Kong, R.Y.C.; Khan, K.M.; Lau, E.T.; Thomas, J.T.; Boot-Handford, R.P.; Grant, M.E.; C
Eur. J. Biochem. 213, 99-111, 1993
A:Title: Intron-exon structure, alternative use of promoter and expression of the mouse
A:Reference number: S31216; MUID:93238750; PMID:8477738
A:Accession: S31216
A:Molecule type: DNA
A:Residues: 1-680 <KON>
A:Cross-references: EMBL:Z21610; NID:949793; PIDN:CAA79736.1; PID:949794
R:Elima, K.; Berola, I.; Rosati, R.; Meteeranta, M.; Garofalo, S.; Peraelae, M.; de Crc
Biochem. J. 289, 247-253, 1993
A:Title: The mouse collagen X gene: complete nucleotide sequence, exon structure and exp
A:Reference number: S28807; MUID:93143676; PMID:8424763
A:Accession: S28807
A:Molecule type: DNA
A:Residues: 1-285, 'A', 287-680 <ELI>
A:Cross-references: EMBL:X67349; NID:950480; PIDN:CAA47763.1; PID:950481
R:Elima, K.; Meteeranta, M.; Kallio, J.; Peraelae, M.; Berola, I.; Garofalo, S.; de Crc
Biochim. Biophys. Acta 1130, 78-80, 1992
A:Title: Specific hybridization probes for mouse alpha-2(X) and alpha-1(X) collagen mRN
A:Reference number: S22215; MUID:92182017; PMID:1543751
A:Accession: S22215
A:Molecule type: mRNA
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 385-450, 'K', 452-627 <ELIA>
A:Cross-references: EMBL:X63013; NID:949795; PIDN:CAA44741.1; PID:949796
R:Acte, S.S.; Olsen, B.R.
Matrix 13, 165-179, 1993
A:Title: Characterization of the mouse type X collagen gene.
A:Accession: S30127; MUID:93261348; PMID:8492743
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-112, 'F', 14-26, 'S', 28-247, 'L', 249-285, 'A', 287-305, 'F', 307-416, 'S', 418-499, 'L
R:Acte, S.S.; Seidlin, M.F.; Hayashi, M.; Olsen, B.R.

```

```

Eur. J. Biochem. 206, 217-224, 1992
A:Title: Cloning of the human and mouse type X collagen genes and mapping of the mouse
A:Reference number: 148299; MUID:92267014; PMID:1587271
A:Accession: 148299
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 52-247, 'L', 249-285, 'A', 287-305, 'F', 307-416, 'S', 418-499, 'L', 501-566, 'C', 568,
A:Cross-references: EMBL:X65121; NID:950482; PIDN:CAA46237.1; PID:9667031
R:Summers, T.A.; Irwin, M.H.; Mayne, R.; Balian, G.
J. Biol. Chem. 263, 581-587, 1988
A:Title: Monoclonal antibodies to type X collagen. Biosynthetic studies using an antibo
A:Reference number: S26397; MUID:88087150; PMID:2826450
A:Accession: S26397
A:Molecule type: Protein
A:Residues: 'SDGYFSQ', 24-26, 'KQ' <SUM>
C:Genetics:
A:Gene: Collagen-1
A:Map position: 10
A:Insertions: 51/3
C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology
C:Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-680/Product: collagen alpha 1(X) chain #status predicted <MAT>
F:553-679/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match
Best Local Similarity 25.7%; Score 168.5; DB 2; Length 680;
Matches 56; Conservative 23; Mismatches 82; Indels 57; Gaps 6;

QY 145 GAGCATFSGYLTVYADADADAPAGPAPPEPRSAFSAARTSLVSDAGGPPHQ----- 199
Db 470 GNPGA--PGPAGIATKGLNPT--GPPGPPPR--GHSBPGLPGPPGPPGQAVMP 523
QY 200 -----PLAFDFEYVNIIGSDPD 215
Db 524 GFIKAGRPRLSGMPLVSNHGVTCMPVSAFTVLISKAYPAVAGAPVDFDITLYNRQOHD 583
QY 216 AAAGVFCRLPGATFFSFTLGKLPKRLSYLKMKNDEVQAMYYDDGASRRMOSQSVW 275
Db 584 PRSGITFCRIPGLYFSYHV-HKGMHWVGLKNGTPM--YYDEYSKGYLDOAGSSAI 641
QY 276 LAIRGDVAWLSDHHDGCGANSNGKYITFGSGLYVP 313
Db 642 MELTENDQWMLQLPNAESNGLYSEYVHSSFGSLVAP 679

RESULT 12
S19018
complement subcomponent C1q chain A precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 20-Aug-1999
C:Accession: S19018
R:Petty, F.; Reid, K.B.M.; Loos, M.
J. Immunol. 147, 3988-3993, 1991
A:Title: Gene expression of the A- and B-chain of mouse C1q in different tissues and th
A:Reference number: S19018; MUID:92043789; PMID:1940381
A:Accession: S19018
A:Molecule type: mRNA
A:Residues: 1-245 <PBT>
A:Cross-references: EMBL:X58861; NID:950226; PIDN:CAA41664.1; PID:950227
A:Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal ho
F:116-243/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match
Best Local Similarity 31.0%; Score 168; DB 2; Length 245;
Matches 53; Conservative 16; Mismatches 70; Indels 32; Gaps 4;

QY 6 LGILGPAACWALGPTPGSSSELRSFAARTPLEGTSEMAVTPPKYVNIIGDPPVAT 65
Db 91 LQDSGPGQGLKGVKNGNIRDPAPAFSAIRQNP--TLGNVVFVKVLTNQESPYQNT 148
QY 66 GQPRCVPGAYPFSF-----TAGKAPKSLSVMLVNRDEVQALAPDEQR 111

```

Db 149 GRFICAVPGFYFNFQVISKWDLCLFKSSGGGPRDSLFSNTNKGFLQVLA----- 202  
QY 112 PGARRASQSAMLTLDYDGTWVLRHGHGPH--VALGAPGATFSGYLVAAD 160  
Db 203 -----GGTVLQLRKDEWVIEKDPKGRITVGTENDSIFSGFLIFPSA 245

RESULT 13  
CGHUID  
collagen alpha 1(X) chain precursor - human  
N/Alternate names: procollagen alpha 1(X) chain  
C/Species: Homo sapiens (man)  
C/Date: 22-Nov-1993 #sequence revision 03-Nov-1995 #text change 22-Jun-1999  
C/Accession: S26396; S30086; S15826; S18249; A43901, 151870; S21856  
R/Reichenberger, B.; Beier, F.; LuValle, P.; Olsen, B.R.; von der Mark, K.; Bertling, W.  
FEBS Lett. 311, 305-310, 1992  
A/Title: Genomic organization and full-length cDNA sequence of human collagen X.  
A/Reference number: S26396; MUID:93012005; PMID:1397333  
A/Accession: S26396  
A/Molecule type: DNA  
A/Residues: 1-680 <RET>  
A/Cross-references: EMBL:X68952; EMBL:X72578; EMBL:X72579; EMBL:X72580; GB:S47714; GB:S4  
R/Apte, S.S.  
submitted to the EMBL Data Library, March 1992  
A/Reference number: S30085  
A/Accession: S30085  
A/Molecule type: DNA  
A/Residues: 1-TPFGVWCVCLL, 52-680 <APT>  
A/Cross-references: EMBL:X65120; NID:G23129  
A/Note: The initial difference is probably due to translation of an intronic sequence  
R/Apte, S.; Mattei, M.G.; Olsen, B.R.  
FEBS Lett. 282, 393-396, 1991  
A/Title: Cloning of human alpha-1(X) collagen DNA and localization of the COL10A1 gene  
A/Reference number: S15826; MUID:91243838; PMID:2037056  
A/Accession: S15826  
A/Molecule type: DNA  
A/Residues: 561-647, 'G', 649-666 <AP2>  
A/Cross-references: EMBL:X58879; NID:G30013; PIDN:CAA41686.1; PID:G30014  
R/Thomas, J.T.; Cresswell, C.J.; Raeh, B.; Nicolai, H.; Jones, T.; Solomon, E.; Grant, M  
Biochem. J. 260, 617-623, 1991  
A/Title: The human collagen X gene. Complete primary translated sequence and chromosomal  
A/Reference number: S18249; MUID:92109659; PMID:1764025  
A/Accession: S18249  
A/Molecule type: DNA  
A/Residues: 1-26, 'T', 28-680 <THO>  
A/Cross-references: EMBL:X60382; NID:G30094; PIDN:CAA42933.1; PID:G30095  
A/Note: the sequence from Fig. 3 is inconsistent with that from Fig. 2 in having 532-Ala  
R/Reichenberger, E.; Aigner, T.; von der Mark, K.; Stoss, H.; Bertling, W.  
Dev. Biol. 148, 562-572, 1991  
A/Title: In situ hybridization studies on the expression of type X collagen in fetal hun  
A/Reference number: A43901; MUID:92077285; PMID:1743401  
A/Accession: A43901  
A/Molecule type: mRNA  
A/Residues: 547-656 <R2>  
A/Cross-references: GB:M74050; GB:D57494; NID:G339884; PIDN:AAA61221.1; PID:G553796  
A/Note: sequence extracted from NCBI backbone (NCBIN:69012, NCBI:69014)  
R/Wallis, G.A.; Raeh, B.; Sweetman, W.A.; Thomas, J.T.; Super, M.; Evans, G.; Grant, M.E  
Am. J. Hum. Genet. 54, 169-178, 1994  
A/Title: Amino acid substitutions of conserved residues in the carboxyl-terminal domain  
pe Schmid.  
A/Reference number: 151870; MUID:94136476; PMID:8304336  
A/Accession: 151870  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 520-597, 'D', 599-680 <WAL>  
A/Cross-references: GB:S68531; NID:G545180; PIDN:AAC60615.1; PID:G545181  
A/Note: mutant sequence from patient with metaphyseal chondrodysplasia type Schmid  
C/Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
ed and subsequently O-glycosylated.  
C/Genetics:  
A/Gene: GDB:COL10A1  
A/Cross-references: GDB:128635; OMIM:120110

A/Map position: 6q21-6q22  
A/Introns: 52/1  
A/Note: A defect in this gene may cause Schmid metaphyseal chondrodysplasia  
C/Complex: type X collagen may be a homotrimer  
C/Function:  
A/Description: structural component of extracellular fibrous polymer specifically and c  
be important for skeletogenesis  
C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
C/Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer; hydroxylysine;  
F/1-18/Domain: signal sequence #status predicted <Sig>  
F/19-680/Product: collagen alpha 1(X) chain #status predicted <NC2>  
F/19-56/Domain: amino-terminal nonhelical #status predicted <NC2>  
F/53-519/Region: interrupted helical  
F/530-680/Domain: amino-terminal nonhelical #status predicted <NC1>  
F/553-679/Domain: complement C1q carboxyl-terminal homology <C1Q>  
F/617/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 9.6%; Score 166.5; DB 1; Length 680;  
Best Local Similarity 25.7%; Pred. No. 1e-05;  
Matches 56; Conservative 22; Mismatches 83; Indels 57; Gaps 6;

QY 145 GARGATFSGYLVYADADADAPARPPAPPPRPSAFSAARTRSVGSADGPRHQ----- 199  
Db 470 GSREP--PGPAGIATKGLNPT--GPPEPPGR--GHSGEPGLPPGPPGPGQAAMPE 523  
QY 200 -----PLAFDTEFVNIGDFD 215  
Db 524 GFIRAGORPGLSGRPLVSANQVYGMPSAFVILSKAPAIQTPIDKILYRQOHYD 583  
QY 216 AAGVFPFRLPGAVFFFTGLPKRLTSLVKMKNDDEVQAMTYDDGASRRREMOSQSV 275  
Db 584 PRGIFTCQIPGIVYFVYH-HVKGHWVWGLKNGTPV-YTYDEVTKYGLDQASGSAI 641  
QY 276 LAIRGDVAWLLSHDHGAGVSNHGYTFESGLVYP 313  
Db 642 IDLTENDQWLDLPNBSNGLYSSEYVHSSFGFLVAP 679

RESULT 14  
S23779  
collagen alpha 1(VIII) chain - mouse  
C/Species: Mus musculus (house mouse)  
C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 16-Jun-2000  
C/Accession: S23779  
R/Muragaki, Y.; Shiota, C.; Inoue, M.; Ooshima, A.; Olsen, B.R.; Minomiya, Y.  
Eur. J. Biochem. 207, 895-902, 1992  
A/Title: Alpha-1(VIII)-collagen gene transcripts encode a short-chain collagen polypept  
A/Reference number: S23779; MUID:92362626; PMID:1499564  
A/Accession: S23779  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-743 <MUR>  
A/Cross-references: EMBL:X66976; NID:G50493; PIDN:CAA47387.1; PID:G1359953  
C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
F/616-742/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 9.6%; Score 165.5; DB 1; Length 743;  
Best Local Similarity 25.8%; Pred. No. 1.4e-05;  
Matches 56; Conservative 23; Mismatches 85; Indels 53; Gaps 6;

QY 136 LHGAPHYALGAPGATFSGYLVYADADAPARPPAPPPRPSAFSAARTRSVGS--DA 192  
Db 540 LHGP-----GKGA-----LPGQGPGLPPGPPGPPAPVPTSPGGEYLPDM 587  
QY 193 GPG-----PHQPLAPTEFVNIGDFD 216  
Db 588 GLGIDGVTPHAYAGKKKGHPAYEMPAFTALTVPPPVGAVKEDKLLYNGRQNP 647  
QY 217 AAGVFPFRLPGAVFFFTGLPKRLTSLVKMKNDDEVQAMTYDDGASRRREMOSQSV 276  
Db 648 OTGIFTCVPGVYFVYH-HCKGNVVALPK-NEMMTYDEVTKYKGFIDQASGSAVL 705  
QY 277 ALRGDVAWLLSHDHGAGVSNHGYTFESGLVYP 313



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd

OM protein - protein search, using sw model

```
Run on:      January 12, 2004, 08:12:09 ; Search time 11.8204 Seconds
              (without alignments)
              1308.910 Million cell updates/sec
```

Title: US-10-085-167-2  
Perfect score: 1726  
Sequence: 1 MLPLLLGLGPACWALGPT .....LVYPDLPAPAPGLGSELL 329

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41:\*

**Pred. No.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1726	100.0	329	1	COT4_HUMAN
2	215	12.5	281	1	CQT1_HUMAN
3	213	12.3	243	1	COT5_HUMAN
4	208	12.1	285	1	COT2_HUMAN
5	200	11.6	278	1	COT6_HUMAN
6	199	11.5	258	1	C1RF_MOUSE
7	196	11.4	244	1	APM1_HUMAN
8	196	11.4	247	1	APM1_MOUSE
9	192	11.1	258	1	C1RF_HUMAN
10	186.5	10.8	255	1	GLIC_MOUSE
11	185.5	10.7	289	1	COT7_HUMAN
12	184	10.7	215	1	HP25_TAVSI
13	182	10.5	251	1	C1O5_HUMAN
14	176.5	10.2	253	1	C1O8_MOUSE
15	173.5	10.1	419	1	COLE_TLEPMA
16	173	10.0	744	1	CA18_RABIT
17	170	9.8	245	1	C1OQ_HUMAN
18	168.5	9.8	680	1	CA1A_MOUSE
19	168	9.7	245	1	C1OQ_MOUSE
20	167.5	9.7	674	1	CA1A_CHICK
21	167	9.7	744	1	CA18_HUMAN
22	166.5	9.6	253	1	C1O8_MOUSE
23	166.5	9.6	680	1	CA1A_MOUSE
24	166.5	9.6	743	1	CA18_HUMAN
25	163.5	9.5	246	1	COT3_HUMAN
26	163.5	9.5	674	1	CA1A_BOVIN
27	163	9.4	224	1	CERL_RAT
28	163	9.4	508	1	OTO1_ONCKE
29	160.5	9.3	246	1	C1OQ_MOUSE
30	158	9.2	193	1	CERB_HUMAN
31	156.5	9.1	193	1	CERB_MOUSE
32	151.5	8.8	245	1	C1OQ_HUMAN
33	150	8.7	166	1	HP20_TAVSI

45	93.5	5.4	686	1	VGH_H_PRV3	Q00660	pseudotaurine
44	94	5.4	1045	1	GNB_CELP1	P26225	echinichia
43	95.5	5.5	495	1	DHAL_ECOLI	P23883	caenorhabdi
42	96.5	5.6	493	1	YKA7_CAEEL	P42458	homo sapien
41	95.5	5.6	497	1	FXD2_HUMAN	O60548	strepptomyc
40	97.5	5.6	476	1	LEU2_STRO	O86514	homo sapien
39	98.5	5.7	265	1	Y176_HUMAN	O14681	mus muscicul
38	135.5	7.9	170	1	CA8_MOUSE	P25318	homo sapien
37	139	8.1	201	1	CERT_HUMAN	O99177	homo sapien
36	139.5	8.1	635	1	CA8_HUMAN	P25067	homo sapien
35	147	8.5	1228	1	ECM_HUMAN	Q13201	homo sapien
34	147.5	8.5	215	1	HP27_TAMSI	O06577	tamias sibi

## ALIGNMENTS

ID	COT4 HUMAN	STANDARD;	PRT;	329 AA.
AC	Q9BXJ3;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Complement-clq tumor necrosis factor-related protein 4 precursor.			
GN	CIQTNP4 OR CTRP4.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Holloway J.L., Lok S.;			
RT	"Homo sapiens complement-clq tumor necrosis factor-related protein.";			
RL	Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.			
CC	-1- SIMILARITY: Contains 2 C1Q domains.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announces">http://www.isb-sib.ch/announces</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; AF329838; AAK17962.1; -			
DR	Genew; HGNC:14346; CIQTNP4.			
DR	InterPro; IPR001073; C1q.			
DR	Pfam; PF00386; C1q; 2.			
DR	SMART; SM00110; C1Q; 2.			
DR	PROSITE; PS01113; C1Q; 2.			
KW	Repeat; Signal.			
FT	SIGNAL	1	16	POTENTIAL.
FT	CHAIN	17	329	COMPLEMENT-C1Q TUMOR NECROSIS FACTOR-
FT				RELATED PROTEIN 4.
FT	DOMAIN	23	159	C1Q 1.
FT	DOMAIN	170	314	C1Q 2.
SQ	SEQUENCE	329 AA;	35265 MW;	331CTDBF26036915 CRC64;
Query Match		100.0%;	Score 1726;	DB 1;
Best Local Similarity		100.0%;	Pred. No. 5.9e-132;	Length 329;
Matches 329;	Conservative	0;	Mismatches 0;	Indels 0;
			Gaps 0;	
QY	1	MPELLIGLIGPAACALGPTTPGSSSELRSAPSAARTTPLEGTSEMAVTFDKYVNI	IGCD	60
DB	1	MPELLIGLIGPAACALGPTTPGSSSELRSAPSAARTTPLEGTSEMAVTFDKYVNI	IGCD	60
QY	61	FDVATGQPCRCRVGAYFFSFTAGKAPKHSIVMLVRRNDEVQALAPDEQRRPARRASQ	120	
DB	61	FDVATGQPCRCRVGAYFFSFTAGKAPKHSIVMLVRRNDEVQALAPDEQRRPARRASQ	120	
QY	121	SAMTQLDYGDYVTLRLHGAAPHYALGAPCATGSGYLVADADAPAGPAPPEPSAFS	180	

Db 121 SANLQIDYDVTWLRHGAHPHALGAPGATFSGYLVYADADADAPARGPAPPEPRSAFS 180  
 Oy 181 AARTSLVGSADAGPGRPHOPLAFDTEFVNIIGDPPAAAGVFCRLPGAYFSTTLGKLR 240  
 Db 181 AARTSLVGSADAGPGRPHOPLAFDTEFVNIIGDPPAAAGVFCRLPGAYFSTTLGKLR 240  
 Oy 241 KTLSTVLMKRNREVOAMITDDGASRRRQSSQSVMLALRGDAVWMLSHDHGCGYASNH 300  
 Db 241 KTLSTVLMKRNREVOAMITDDGASRRRQSSQSVMLALRGDAVWMLSHDHGCGYASNH 300  
 Oy 301 GKXITPSGLVYFDLPAAPPGAGSELL 329  
 Db 301 GKXITPSGLVYFDLPAAPPGAGSELL 329

RESULT 2  
 COT1\_HUMAN STANDARD; PRT; 281 AA.  
 ID Q9BXJ1; Q96NF2; Q9GZK4;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Complement-c1q tumor necrosis factor-related protein 1 precursor  
 (c1q protein coupled receptor interacting protein) (GIP).  
 GN C1QTNF1 OR CTRP1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Inamatori G., Le Guill C., Whang I., Birnbaumer M.;  
 RT "GIP, a putative GPCR interacting protein.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Sheppard P.O.;  
 RT "Homo sapiens complement-c1q tumor necrosis factor-related protein.";  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP TISSUE-SKIN;  
 RC MEDLINE=2238257; PubMed=12477932;  
 RX Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marnett K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Rana S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hilyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [4]  
 RP SEQUENCE OF 83-281 FROM N.A.  
 RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,  
 RA Hotta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J., Chiba Y.,  
 RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,  
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,  
 RA Yamamoto K., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,  
 RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,  
 RA Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
 RA Suzuki Y., Sugano S., Nagahara K., Masuno Y., Nagai K., Isegaki T.,

RT "NEDO human cDNA sequencing project.";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Contains 1 collagenous domain.  
 CC -1- SIMILARITY: Contains 1 C1Q domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AJ272138; CAC20425.1; ALT INIT.  
 DR EMBL; AF232905; AAG44303.1; ALT INIT.  
 DR EMBL; AF329840; AAK17964.1; -  
 DR EMBL; BC021553; AAH21553.1; -  
 DR EMBL; AK055541; BAB70947.1; -  
 DR Genew; HGNC:14324; C1QTNF1.  
 DR InterPro; IPR001073; C1Q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1Q; 1.  
 DR Pfam; PF01391; Collagen; 1.  
 DR SMART; SM00110; C1Q; 1.  
 DR PROSITE; PS01113; C1Q; FALSE\_NEG.  
 KW Collagen; signal.  
 FT SIGNAL 1 25  
 FT CHAIN 26 281  
 FT  
 FT DOMAIN 99 140 POTENTIAL.  
 FT DOMAIN 147 281 COMPLEMENT-C1Q TUMOR NECROSIS FACTOR-  
 FT CONFLICT 22 22 C1Q. RELATED PROTEIN 1.  
 FT CONFLICT 241 241 C1Q. COLLAGEN-LIKE.  
 FT CONFLICT 241 241 L -> P (IN REF. 1).  
 FT CONFLICT 241 241 R -> Q (IN REF. 1).  
 SQ SEQUENCE 281 AA; 31743 MW; 49E248CB8ACFE7C CRC64;

Query Match 12.5%; Score 215; DB 1; Length 281;  
 Best Local Similarity 30.2%; Pred. No. 2.7e-10;  
 Matches 77; Conservative 26; Mismatches 86; Indels 66; Gaps 11;

Oy 69 RCRVPGAYFSTFGARPHKSHSIVLVRNDEVOALAFDQRPRGARRAASQSMQLDLY 128  
 Db 75 RCDPFGTSMYPARA--VPQINTITLKEKCDR-----GDRG-----LOGKY 113  
 Oy 129 GDT--VWLRHGAAP---HYALGAPGATFSGYLVYADADADAPARGPAPPEPRSAFS 183  
 Db 114 GKIGSAGARGHTGPKGKSGMGPGRCKSH--YA-----AFSVGR 152  
 Oy 184 TRSLVGSADAGPGRH-----QPLAFDTEFVNIIGDPPAAAGVFCRLPGAYFSTTLGK 238  
 Db 153 KK-----PMHSNHYQTIVFDTEFVNLDFHFMFTGKFCYVGLYFSLNVHTW 202  
 Oy 239 PRKTLSTVLMKRNREVOAMITDDGASRRRQSSQSVMLALRGDAVWMLSHDHGCGYAS 298  
 Db 203 NQKSTVYHMKNEEVEVILPRAVG--DRSITQSLSLWLEIREDDQVWRLYKGERNAIF 260  
 Oy 299 NH--GKXITPSGLV 311  
 Db 261 SEEDPTITTSGLIV 275

RESULT 3  
 COT5\_HUMAN STANDARD; PRT; 243 AA.  
 ID Q9BXJ0; Q9UFY4;  
 AC Q9BXJ0; Q9UFY4;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Complement-c1q tumor necrosis factor-related protein 5 precursor.  
 GN C1QTNF5 OR CTRP5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;

[1]  
 RN SEQUENCE FROM N.A.  
 RA Sheppard P.O., Humes J.M.;  
 RT "Homo sapiens complement-c1q tumor necrosis factor-related protein.";  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 25-243 FROM N.A.  
 RC TISSUE=Uterus;  
 RA Ottenwelder B., Obermaier B., Mewes H.-W., Gassenhuber J.,  
 RA Wiemann S.;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Contains 1 collagenous domain.  
 CC -1- SIMILARITY: Contains 1 C1Q domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF329841; AAK17965.1; -  
 DR EMBL; AL110261; CAB53702.1; -  
 DR PIR; T14782; T14782  
 DR Genew; HGNC:14344; C1QTNF5.  
 DR InterPro; IPR001073; C1Q.  
 DR Pfam; PF00386; C1Q; 1.  
 DR Pfam; PF00386; C1Q; 1.  
 DR PRINTS; PRO0007; COLLAGEN; 1.  
 DR SMART; SM00110; C1Q; 1.  
 DR PROSITE; PS01113; C1Q; FALSE\_NEG.  
 DR KEGG; collagen; signal.  
 FT CHAIN 1 15 POTENTIAL.  
 FT SIGNAL 16 243 COMPLEMENT-C1Q TUMOR NECROSIS FACTOR-  
 FT DOMAIN 30 95 RELATED PROTEIN 5.  
 FT DOMAIN 97 243 COLLAGEN-LIKE.  
 FT DOMAIN 97 243 C1Q.  
 SQ SEQUENCE 243 AA; 25298 MW; 7CCDA65CDA7EB784 CRC64;  
 Query Match 12.3%; Score 213; DB 1; Length 243;  
 Best Local Similarity 34.9%; Pred. No. 3.3e-10;  
 Matches 66; Conservative 18; Mismatches 83; Indels 22; Gaps 6;  
 QY 138 GAPHYALGAPGATSG--YVYVADAD---ADAPAGPAPAP-----EPSSAFSAARTR 185  
 DB 57 GAP-----GAPGKEGEGRPGLPGRGDPGPRGEGAPGPGPAGECVPPRSASAKRSE 112  
 QY 186 SLVSGDAGPGRHOPPLAFDTEFVNIIGDFFDAAGVFRCLGAYFSTFLGKLPKRLTSV 245  
 DB 113 SRV-----PPPSDAPLPFRDLVNEQGHYDAVTGKTCQVGVYVFA-VNATVYRASLQF 166  
 QY 246 KLKMNREVMQAMITDDGASRRKREKOSQSVMLALRGDAVWLISHDHGYGAYSHGKYIT 305  
 DB 167 DLVNGSISLAFQFGWPKPDSLSCGAWRLPEPDQVWVGVGVIYIGIYIKTST 226  
 QY 306 FSGFLVYPD 314  
 DB 227 FSGFLVYSD 235  
 RESULT 4  
 COT2 HUMAN  
 ID\_COT2 HUMAN STANDARD; PRT: 285 AA.  
 AC OPHX3;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Complement-c1q tumor necrosis factor-related protein 2 precursor.  
 GN C1QTNF2 OR CTRP2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Piddington C.S., Bishop P.;  
 RT "Homo sapiens complement-c1q tumor necrosis factor-related protein.";  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Muscle;  
 RX MEDLINE=22388257; Pubmed=12477932;  
 RA Krausberg R.L., Faingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
 RA Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Saperstein M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalek A., Smalish D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- SIMILARITY: Contains 1 collagenous domain.  
 CC -1- SIMILARITY: Contains 1 C1Q domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF329836; AAK17960.1; -  
 DR EMBL; BC011699; AAK11699.1; -  
 DR Genew; HGNC:14325; C1QTNF2.  
 DR InterPro; IPR001073; C1Q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1Q; 1.  
 DR Pfam; PF01391; Collagen; 2.  
 DR PRINTS; PRO0007; COLLAGEN; 1.  
 DR SMART; SM00110; C1Q; 1.  
 DR PROSITE; PS01113; C1Q; 1.  
 DR KEGG; collagen; signal.  
 FT SIGNAL 1 15 POTENTIAL.  
 FT CHAIN 16 285 COMPLEMENT-C1Q TUMOR NECROSIS FACTOR-  
 FT DOMAIN 40 141 RELATED PROTEIN 2.  
 FT DOMAIN 143 285 COLLAGEN-LIKE.  
 FT DOMAIN 143 285 C1Q.  
 SQ SEQUENCE 285 AA; 29952 MW; 7E31FP9868D4EDFA CRC64;  
 Query Match 12.1%; Score 208; DB 1; Length 285;  
 Best Local Similarity 33.3%; Pred. No. 1e-09;  
 Matches 55; Conservative 31; Mismatches 49; Indels 30; Gaps 8;  
 QY 7 GLIGPAACVALGTPPGSSSELSAFAAA--RTPLLEGISSEMAVTDKYYVNIIGCPDVA 64  
 DB 136 GLPGPSC-----GSGHTSAPSVAVTKSYPRE--RLPIKDKLIMEGHHYNAS 183  
 QY 65 TQGFRCRVGCAVFFSTAGAPKHSLSVWLVNRDQVOLAFAPEQGRPGARRASQSAMU 124  
 DB 184 SGKFCVGVCEIYFTYDITLA-NKHLATGLVHN-GQYRIRTPD--ANTGNHVAAGSSTIL 239  
 QY 125 QLDYGDVTWLRLHGA-----PHYALGAPGATFSGYLTVADAD 161







CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D45371; BA08327.1; -  
 DR EMBL; AB012165; BA086716.1; -  
 DR EMBL; AB012164; BA086716.1; JOINED.  
 DR EMBL; AJ131460; CAB52413.1; -  
 DR EMBL; AJ131461; CAB52413.1; JOINED.  
 DR PIR; JC4708; JC4708.  
 DR MIM; 605441; -  
 DR GO; 0006091; P-energy pathways; TNS.  
 DR InterPro; IPR01073; C1q.  
 DR InterPro; IPR00087; Collagen.  
 DR Pfam; PF00386; C1q; 1.  
 DR Pfam; PF01391; collagen; 1.  
 DR PRINTS; PR00007; COMPLEMENTC1Q.  
 DR Prodom; PD00007; C1q\_helix; 1.  
 DR SMART; SM00110; C1Q; 1.  
 DR PROSITE; PS01113; C1Q; 1.  
 KW Hormone; Collagen; Signal; Repeat; Hydroxylation; Plasma;  
 KW Polymorphism; Disease mutation; Obesity; Diabetes mellitus.  
 FT SIGNAL 1 14  
 FT CHAIN 15 244  
 FT DOMAIN 42 107  
 FT DISULFID 108 244  
 FT MOD\_RES 36 244  
 FT MOD\_RES 44 44  
 FT MOD\_RES 47 44  
 FT MOD\_RES 53 53  
 FT MOD\_RES 62 62  
 FT MOD\_RES 71 71  
 FT MOD\_RES 76 76  
 FT MOD\_RES 86 86  
 FT MOD\_RES 95 95  
 FT MOD\_RES 104 104  
 FT VARIANT 84 84  
 FT VARIANT 112 112  
 FT VARIANT 117 117  
 FT VARIANT 164 164  
 FT VARIANT 221 221  
 FT VARIANT 241 241  
 FT SEQUENCE 244 AA; 26414 MW; 6408C6C1204B1018 CRC64;  
 Query Match 11.4%; Score 196; DB 1; Length 244;  
 Best Local Similarity 29.7%; Pred. No. 7.7e-09;  
 Matches 54; Conservative 31; Mismatches 61; Indels 36; Gaps 8;  
 QY GILGP-----AACWALGP-----TPGQSSELSRFAFAARTTLEGTSEAV 48  
 DB GILGPKGIDIGTGVGAEGRGFGPIQGRKGEKSGAVYRASFVGLTETVT-IPNMP 130  
 QY 49 TEDKYVYVIGSDFDVATGFRGVRGAYFEFTACKRPH-----KSLSLVLRNDDVQA 103  
 DB 131 RTTKLFYVQNHVDSGTGFHCNIPGLYYFAY-----HTVYMDVVRKSLFK-KDKML 183  
 QY 104 LAFDEQRGARRAASQSMLOLDYDPTVWLRHGAAPH-----YALGAPATFSGYLVAD 159  
 DB 164 FTYDQGENNVQD-ASGSVLHLHLEVGDDQVWLQVYGEGERNGXADNDNDSTTGFLIHD 242  
 QY 160 AD 161  
 DB 243 TN 244

RESULT 8  
 APML\_MOUSE

ID APML\_MOUSE STANDARD; PRT; 247 AA.  
 AC Q60994; Q62400; Q9DC68;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, last sequence update)  
 DT 15-SEP-2003 (Rel. 42, last annotation update)  
 DE Adiponectin precursor (30 kDa adipocyte complement-related protein)  
 DE (ACRP30) (Adipocyte specific protein Adipoq).  
 GN APML OR ACRP30 OR ADIPOQ.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCBI\_TaxID=10090;  
 RN [1]  
 RP TISSUE=Adipocyte;  
 RC MEDLINE=96070757; PubMed=7592907;  
 RA Scherer P.E., Williams S., Fogliano M., Baldini G., Lodish H.F.;  
 RT "A novel serum protein similar to C1q, produced exclusively in  
 adipocytes.";  
 RL J. Biol. Chem. 270:26746-26749(1995).  
 RN [2]  
 RP TISSUE=Fibroblast;  
 RC MEDLINE=96209999; PubMed=8631877;  
 RA Hu E., Liang P., Spiegelman B.M.;  
 RT "Adipoq is a novel adipose-specific gene dysregulated in obesity.";  
 RL J. Biol. Chem. 271:10697-10703(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC PubMed=11162643;  
 RA Das K., Lin Y., Widen E., Zhang Y., Scherer P.E.;  
 RT "Chromosomal localization, expression pattern, and promoter analysis  
 of the mouse gene encoding adipocyte-specific secretory protein  
 Acrp30.";  
 RL Biochem. Biophys. Res. Commun. 280:1120-1129(2001).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Heart;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 RA Aizawa K., Iwata M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
 RA Kuehl P., Lewis S., Matsuoka Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staab J.F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Botfield D., Bjurung N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazarelli J., Momberears P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyshak-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [5]  
 RP FUNCTION.  
 RX MEDLINE=21372498; PubMed=11479627;  
 RA Yamauchi T., Kamon J., Waki H., Terachi Y., Kubota N., Hara K.,  
 RA Mori Y., Ide T., Murakami K., Tsuboyama-Kaoaka N., Ezaki O.,  
 RA Akanuma Y., Gavrilova O., Vinson C., Reitman M.L., Kagechika H.,  
 RA Shudo K., Yoda M., Nakano Y., Tobe K., Nagai R., Kimura S., Tomita M.,  
 RA Froguel P., Kadowaki T.;  
 RT "The fat-derived hormone adiponectin reverses insulin resistance  
 associated with both lipodystrophy and obesity.";  
 RL Nat. Med. 7:941-946(2001).  
 RN [6]  
 RP FUNCTION.  
 RX MEDLINE=21372499; PubMed=11479628;



CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AF095154; AAC64186.1; -

DR EMBL; AF410771; AAC65248.1; -

DR EMBL; BC008798; AAH08798.1; -

DR GO; GO:0007626; P:locomotory behavior; NAS.

DR InterPro; IPR001073; C1Q.

DR InterPro; IPR000087; Collagen.

DR Pfam; PF00386; C1q; 1.

DR Pfam; PF01391; Collagen; 1.

DR PRINTS; PR00007; COMPLEMENTC1Q.

DR SMART; SM00110; C1Q; 1.

DR PROSITE; PS01113; C1Q; 1.

DR Collagen; Signal.

FT SIGNAL 1 16 POTENTIAL.

FT CHAIN 17 258 C1Q-RELATED FACTOR.

FT DOMAIN 67 115 COLLAGEN-LIKE.

FT DOMAIN 123 258 C1Q.

FT SEQUENCE 258 AA; 26452 MW; 52C51CDF59CAE2BF CRC64;

Query Match 11.1%; Score 192; DB 1; Length 258;  
Best Local Similarity 27.3%; Pred. No. 1.7e-08;  
Matches 67; Conservative 26; Mismatches 88; Indels 64; Gaps 9;

CC -----

CC 110 RRGARRAASQAMQLDYPDTWMLRHGAPHYAL-----GAPGATFSGLYVADADADA.164

CC 38 RGFPGAG-----ARTGQDALSRQSGAPRPSLTVGPGKRGRT-----GKP 78

CC 165 PARGPAPPEPRSAFSAARTSLVGSADAPG-----PR-----197

CC 79 GPPGPPDPGPPGPPGPEKPEKPPGPGPLPGAGSGAISTATYTTVPVAFYAGLKN 138

CC 198 -----HQPLAPDFEFNIGDDPDAAGVRCRLPGAFYFSFTLGKLP-----KTLSSVKLMK 249

CC 139 PHEGEVLFKFDVVTNLNNYDAAGKFTCNIPGTFYFVHV--LMRGDGTSMWADLCK 196

CC 250 NRDEYQAMLYDDGASRRREMOSQSYMLALRRGDVAWLSHDHGAGYVSNHGKYYTFSGF 309

CC 197 N-GQVRASAIADQADQNDVYASNSVYLHLHDAGDEVFIKLDGSKAHGNSN--KYSTFSGF 253

CC 310 LVYPD 314

CC 254 IYSD 258

CC -----

CC RESULT 10

CC GLIC\_MOUSE STANDARD; PRT: 255 AA.

CC AC Q9ESN4;

CC DT 16-OCT-2001 (Rel. 40, Created)

CC DT 16-OCT-2001 (Rel. 40, Last sequence update)

CC DT 28-FEB-2003 (Rel. 41, Last annotation update)

CC DE Glacolin precursor (C1q-like protein).

CC GN C1Q.

CC OS Mus musculus (Mouse).

CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

CC OX NCBI\_TaxID=10090;

CC RN NCBI\_TaxID=10090;

CC RP SEQUENCE FROM N.A.

CC RX MEDLINE=20428709; PubMed=1086216;

CC RA Koide T., Aso A., Yoshizaki T., Nagata K.

CC RT "Conformational requirements of collagenous peptides for recognition by the chaperone protein HSP47."

CC RL J. Biol. Chem. 275:27957-27963(2000).

CC CC -1- TISSUE SPECIFICITY: EXPRESSED IN GLIAL CELLS.

CC CC -1- SIMILARITY: Contains 1 collagenous domain.

CC CC -1- SIMILARITY: Contains 1 C1Q domain.

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AB044560; BAB15806.1; -

DR MG; MG1:2387350; C1qL.

DR GO; GO:0005515; F:protein binding activity; IPI.

DR InterPro; IPR001073; C1Q.

DR InterPro; IPR000087; collagen.

DR Pfam; PF00386; C1q; 1.

DR Pfam; PF01391; Collagen; 1.

DR PRINTS; PR00007; COMPLEMENTC1Q.

DR SMART; SM00110; C1Q; 1.

DR PROSITE; PS01113; C1Q; 1.

DR Collagen; Signal.

FT SIGNAL 1 20 POTENTIAL.

FT CHAIN 21 255 GLIACOLIN.

FT DOMAIN 61 111 COLLAGEN-LIKE.

FT DOMAIN 120 255 C1Q.

FT SEQUENCE 255 AA; 26687 MW; 529FBAFAB2191BC1 CRC64;

Query Match 10.8%; Score 186.5; DB 1; Length 255;  
Best Local Similarity 26.3%; Pred. No. 4.7e-08;  
Matches 57; Conservative 23; Mismatches 86; Indels 51; Gaps 6;

CC -----

CC 136 LHGAPHYALGAPATFSGLYVADADADAPAPGPAPPEPRSAFSAARTSLVGSADAPG 195

CC 52 MQSLPFTIQPKGE-----AGRPKAGRGPRGPPGPPGPEKPEPRGGLPG 103

CC 196 PRHP-----LAFDFEFNIGDDPDAAGV 221

CC 104 PPGAPGALNAGALSAATYSTVPKIAFYAGLKRQHEGEVLFKFDVVTNLGNHYDPTTKF 163

CC 222 RCRPGAFYFSFTLGKLP-----KTLSSVKLMNRDEYQAMLYDDGASRRREMOSQSYMLA 277

CC 164 TCSIPGIFYFVHV--LMRGDGTSMWADLCKN--NQRASAIADQADQNDVYASNSVYLH 220

CC 278 LRRGDVAWLSHDHGAGYVSNHGKYYTFSGFLYPD 314

CC 221 LEPGEVYIKLDGSKAHG--GNNNKYGTFSGFIYAD 255

CC -----

CC RESULT 11

CC COT7\_HUMAN STANDARD; PRT: 289 AA.

CC AC Q9BXO2;

CC DT 28-FEB-2003 (Rel. 41, Created)

CC DT 28-FEB-2003 (Rel. 41, Last sequence update)

CC DT 15-SEP-2003 (Rel. 42, Last annotation update)

CC DE Complement-c1q tumor necrosis factor-related protein 7 precursor.

CC GN C1QTNF7 OR CTRP7.

CC OS Homo sapiens (Human).

CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

CC OX NCBI\_TaxID=9606;

CC RN NCBI\_TaxID=9606;

CC RP SEQUENCE FROM N.A.

CC RA Piddington C.S., Sheppard P.O., Bishop P., Lasser G.W.

CC RT "Homo sapiens complement-c1q tumor necrosis factor-related protein.";

CC RL Submitted (DDB-2000) to the EMBL/GenBank/DBJ databases.

CC RN [2]

CC RP SEQUENCE FROM N.A.

CC RC TISUE-Testis;

CC RX MEDLINE=22386257; PubMed=12477932;

CC RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Alekshun S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares W.B., Bonaldo M.F., Casavant T.L., Schetz T.E., Brownstein M.J., Ueda T.B., Tomihata S., Carninci P., Prange C., Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,

RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Halyk S.W.,  
 RA Villalón D.K., Wozny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
 RA Schercher A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RL human and mouse cDNA sequences.";  
 CC Sci. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 CC -1- SIMILARITY: Contains 1 ClQ domain.  
 CC -1- SIMILARITY: Contains 1 ClQ domain.  
 CC -1- SIMILARITY: Contains 1 ClQ domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: AF329839; AAK17963.1; -  
 DR EMBL: BC022187; AAH22187.1; -  
 DR Genew; HNC:14342; C1QNF7.  
 DR InterPro: IPR001073; ClQ.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF00386; ClQ; 1.  
 DR Pfam: PF01391; Collagen; 2.  
 DR PRINTS: PR00007; COMPLEMENTC1Q.  
 DR SMART: SM00110; ClQ; 1.  
 DR PROSITE: PS01113; ClQ; 1.  
 DR Collagen; Signal.  
 FT CHAIN 1 16 POTENTIAL.  
 FT SIGNAL 17 289 COMPLEMENT-C1Q TUMOR NECROSIS FACTOR-  
 FT DOMAIN 38 139 RELATED PROTEIN 7.  
 FT DOMAIN 141 276 COLLAGEN-LIKE.  
 FT SEQUENCE 289 AA; 30683 MW; A61609FF86D26946 CRC64;  
 SQ  
 Query Match 10.7%; Score 185.5; DB 1; Length 289;  
 Best Local Similarity 33.5%; Pred. No. 6.6e-08;  
 Matches 60; Conservative 22; Mismatches 60; Indels 37; Gaps 9;  
 QY 7 GILGPACWALGPTPGP-----GSSELRSAFSAARTT--PLEGTSEM 46  
 Db 113 GEVGP-----IGP-PGPKGDGEGDGLPGVCRGSLVLSAIVSGITTSPEE--RL 163  
 QY 47 AVTFDKYVNIIGDQDVATGQFRCRVPGAYFFSFTAGKAPKSLVLMVNRDEVQALAF 106  
 Db 164 PLIFNKVLFNENGEHNPATGKFCAPGIYFESYDITLA-NKHLAIGLVNH-GQYRIKTF 221  
 QY 107 DEQRPRGARRAASOSAMQLDYGPTVMRLHGAHYAL---GAPGATFSGVLYAADAD 161  
 Db 222 D-ANTGNHDVASSSTVIYILOPEDEWMLIEFTDONGLFSDPGWADSLFSGFLLYVDPD 278

RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=93180798; Pubmed=8441393;  
 RA Takamatsu N., Ohba K., Kondo J., Kondo N., Shiba T.,  
 RT "Hibernation-associated gene regulation of plasma proteins with a  
 RL collagen-like domain in mammalian hibernators.";  
 RL Mol. Cell. Biol. 13:1516-1521 (1993).  
 RM [2]  
 RP SEQUENCE OF 29-62; 84-130; 172-183; 187-192 AND 201-215.  
 RC TISSUE=Plasma;  
 RX MEDLINE=92112696; Pubmed=1730610;  
 RA Kondo N., Kondo J.;  
 RT "Identification of novel blood proteins specific for mammalian  
 RT hibernation.";  
 RL J. Biol. Chem. 267:473-478 (1992).  
 CC -1- FUNCTION: PLASMA PROTEINS HP-20, HP-25, HP-27 AND HP-55 FORM A  
 CC 140 kDa COMPLEX VIA DISULFIDE BONDS IN THE PLASMA AND ARE  
 CC HIBERNATION SPECIFIC.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: PLASMA; SYNTHESIZED IN THE LIVER.  
 CC -1- DEVELOPMENTAL STAGE: THE PROTEIN COMPLEX DISAPPEARS FROM THE  
 CC PLASMA AT ONSET OF HIBERNATION AND REAPPEARS AS HIBERNATION  
 CC CEASES.  
 CC -1- SIMILARITY: Contains 1 collagenous domain.  
 CC -1- SIMILARITY: Contains 1 ClQ domain.  
 CC -1- SIMILARITY: Contains 1 ClQ domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: D12975; BA02352.1; -  
 DR PIR: B48150; B48150.  
 DR InterPro: IPR001073; ClQ.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF00386; ClQ; 1.  
 DR Pfam: PF01391; Collagen; 1.  
 DR PRINTS: PR00007; COMPLEMENTC1Q.  
 DR SMART: SM00110; ClQ; 1.  
 DR PROSITE: PS01113; ClQ; 1.  
 DR Signal; Collagen; Glycoprotein; Plasma; Multigene family.  
 FT SIGNAL 1 28  
 FT CHAIN 1 215 HIBERNATION-ASSOCIATED PLASMA PROTEIN  
 FT HP-25  
 FT DOMAIN 40 81 COLLAGEN-LIKE.  
 FT DOMAIN 167 215 ClQ.  
 FT CARBOHYD 167 167 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 215 AA; 22664 MW; AFE03206917EA530 CRC64;  
 Query Match 10.7%; Score 184; DB 1; Length 215;  
 Best Local Similarity 30.4%; Pred. No. 6.2e-08;  
 Matches 62; Conservative 22; Mismatches 52; Indels 68; Gaps 11;  
 QY 138 GAPHYALGAPGATFSGVLYAADADAPAGPPAPP-----EPSS 177  
 Db 49 GIPGFP-GAPGAL-----GPGPEPVPIPIPGPGPGVCKSSPKS 90  
 QY 178 AFSARTSLVGSADAGPRHOPLADEFTFVNIIGDFAAGVFCRLPGAYFFSFTLTK 237  
 Db 91 AFAVKL-----SERPREP-FQPIVFEALYNOGHPNMTAGERSCVLPVYNGFDI-R 142  
 QY 238 LPKRTLSVKMKRDEVOAMITVDDGASRRREMSSQ-----SYMLALRGDAVWLSH 289  
 Db 143 LFQSSVIRLM-RDGIQV-----REXEAQANDSYKIAMSVTIALGDKVWLESK 192  
 QY 290 DHQGYGAYNHG-KYITFSGLVY 312  
 Db 193 LK---GHESEKGIITHIVFGYLLX 213

```

13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
10
```

RN		[7]	REVIEW OF C1Q DEFICIENCY.
RP			MEDLINE=98450587; PubMed=9777412;
RA	Petry F.;		"Molecular basis of hereditary C1q deficiency.";
RT	Immunobiology 199;286-294(1998)		
RL		-I-	FUNCTION: C1Q ASSOCIATES WITH THE PROENZYMES C1R AND C1S TO YIELD C1, THE FIRST COMPONENT OF THE SERUM COMPLEMENT SYSTEM. THE COLLAGEN-LIKE REGIONS OF C1Q INTERACT WITH THE CA(2+)-DEPENDENT C1R(2)C1S(2) PROENZYMIC COMPLEX, AND EFFICIENT ACTIVATION OF C1 TAKES PLACE ON INTERACTION OF THE GLOBULAR HEADS OF C1Q WITH THE FC REGIONS OF IGG OR IGM ANTIBODY PRESENT IN IMMUNE COMPLEXES.
CC		-I-	SUBUNIT: C1 IS A CALCIUM-DEPENDENT TRIMOLECULAR COMPLEX OF C1Q, R AND S IN THE MOLAR RATIO OF 1:1:2. C1Q SUBCOMPONENT IS COMPOSED OF NINE SUBUNITS, SIX OF WHICH ARE DISULFIDE-LINKED DIMERS OF THE A AND B CHAINS, AND THREE OF WHICH ARE DISULFIDE-LINKED DIMERS OF THE C CHAIN.
CC		-I-	PTM: O-linked glycans consist of Glc-Gal disaccharides bound to the oxygen atom of post-translationally added hydroxyl groups.
CC		-I-	DISEASE: Defects in C1QB are a cause of C1q deficiency [MM:20570]. It is a rare genetic disorder which is associated with recurrent infections and a high prevalence of lupus erythematosus-like symptoms. It is characterized by a loss of activation of the complement classical pathway.
CC		-I-	SIMILARITY: Contains 1 collagenous domain.
CC		-I-	SIMILARITY: Contains 1 C1q domain.
CC			-----
CC			This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch</a> ).
CC			-----
DR	EMBL, X03084; CAA26880.1; .		
DR	EMBL, BC008963; AAH08963.1; ALT_INIT.		
DR	EMBL, M36278; AAC41692.1; .		
DR	Genew; HGNC:1242; C1QB.		
DR	MIM: 120570; .		
DR	GO; GO:0005602; C:complement component C1q complex; TAS.		
DR	GO; GO:0003181; F:complement activity activity; TAS.		
DR	InterPro; IPR001073; C1q.		
DR	InterPro; IPR000087; Collagen.		
DR	Pfam; PF00386; C1q; 1.		
DR	Pfam; PF01391; collagen; 1.		
DR	PRINTS; PR00007; COMPLEMENTC1Q.		
DR	SMART; SM00110; C1Q; 1.		
KW	PROSITE; PS01113; C1Q; 1.		
DR	Complement pathway; Plasma; Hydroxylation; Glycoprotein; Collagen; Repeat; Signal; Disease mutation; Pyroglutamate carboxylic acid.		
FT	SIGNAL	1	25
FT	CHAIN	26	251
FT	DOMAIN	29	112
FT	DOMAIN	113	251
FT	MOD_RES	26	26
FT	DISULFD	29	29
FT	MOD_RES	33	33
FT	MOD_RES	36	36
FT	MOD_RES	39	39
FT	MOD_RES	42	42
FT	MOD_RES	51	51
FT	MOD_RES	54	54
FT	MOD_RES	57	57
FT	CARBOHYD	57	57
FT	MOD_RES	60	60
FT	CARBOHYD	60	60
FT	MOD_RES	63	63
FT	MOD_RES	75	75
FT	MOD_RES	81	81
CC			COMPLEMENT C1Q SUBCOMPONENT, B CHAIN.
CC			COLLAGEN-LIKE.
CC			C1Q.
CC			PYROGLUTAMATE CARBOXYLIC ACID.
CC			INTERCHAIN (WITH C-26 IN CHAIN A).
CC			HYDROXYLATION.
CC			HYDROXYLATION.
CC			HYDROXYLATION.
CC			HYDROXYLATION.
CC			HYDROXYLATION.
CC			O-LINKED (GAL. . .).
CC			HYDROXYLATION.
CC			O-LINKED (GAL. . .).
CC			HYDROXYLATION.
CC			HYDROXYLATION.
CC			HYDROXYLATION.
CC			HYDROXYLATION.

FT MOD\_RES 84 84 HYDROXYLATION.  
 FT MOD\_RES 90 90 HYDROXYLATION.  
 FT MOD\_RES 96 96 HYDROXYLATION.  
 FT CARBOHYD 96 96 O-LINKED (GAL. .).  
 FT MOD\_RES 99 99 HYDROXYLATION.  
 FT MOD\_RES 102 102 HYDROXYLATION.  
 FT MOD\_RES 105 105 HYDROXYLATION.  
 FT MOD\_RES 108 108 HYDROXYLATION.  
 FT CARBOHYD 108 108 O-LINKED (GAL. .).  
 FT VARIANT 40 40 G-> D (in C1Q deficiency).  
 FT CONFLICT 26 26 /FTID=VAR 008541.  
 FT CONFLICT 83 83 Q-> E (IN REF. 3).  
 FT CONFLICT 98 98 N-> D (IN REF. 3).  
 SQ SEQUENCE 251 AA; 26459 MW; 78C575267A0EF7 CRC64;  
 Query Match 10.5%; Score 182; DB 1; Length 251;  
 Best Local Similarity 29.1%; Pred. No. 1.1e-07;  
 Matches 48; Conservative 32; Mismatches 61; Indels 24; Gaps 5;  
 QY 165 PARPPAPPEP-----RSAFSAARSLVSGSDAGPGRHOPLAFTDFVNI 211  
 DB 95 PKGPGAGAGAPGPKSGSDYAKATQKIAFSATRTINV-----PLRQDTIRFDHYITMMN 148  
 QY 212 GDPDAAGVFRGRLPGAFPFSPFLTKLPRKTLSTYKLMKRDVEQAMT-YDGCASRRREQ 270  
 DB 149 NNYEPRSGKFTCKPGGLYFTYHASS--RGNLCVLMKRGERRAOKVTFCDYAVNTFOVT 206  
 QY 271 SOSVWLALRGDAVWVLSHDHGYGAYSNHGKTYTFSGFLYPPDL 315  
 DB 207 TGGVVLKLEGGENVFLQATDKN--SLGCMEGANSIFSGFLYPPDM 249  
 RESULT 14  
 ID C1QB RAT STANDARD; PRT; 253 AA.  
 AC P31721;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Complement C1q subcomponent, B chain precursor.  
 GN C1QB.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Spleen;  
 RX MEDLINE=96062319; PubMed=7594503;  
 RA Schwaebel W., Schaefer M.K.-H., Petry F., Fink T., Knebel D.,  
 RA Weihe E., Loope M.;  
 RT "Pollicular dendritic cells, interdigitating cells, and cells of the  
 RT monocyte-macrophage lineage are the C1q-producing sources in the  
 RT spleen. Identification of specific cell types by in situ  
 RT hybridization and immunohistochemical analysis.";  
 RL J. Immunol. 155:4971-4978 (1995).  
 RN [2]  
 RP SEQUENCE OF 71-79 AND 141-146.  
 RX MEDLINE=93218657; PubMed=8464426;  
 RA Wing M.G., Seilly D.J., Bridgman D.U., Harrison R.A.;  
 RL "Rapid isolation and biochemical characterization of rat C1 and C1q";  
 RL Mol. Immunol. 30:433-440 (1993).  
 CC -1- FUNCTION: C1Q ASSOCIATES WITH THE PROENZYMES C1R AND C1S TO YIELD  
 CC C1, THE FIRST COMPONENT OF THE SERUM COMPLEMENT SYSTEM. THE  
 CC COLLAGEN-LIKE REGIONS OF C1Q INTERACT WITH THE CA(2+)-DEPENDENT  
 CC C1R(2)C1S(2) PROENZYME COMPLEX, AND EFFICIENT ACTIVATION OF C1  
 CC TAKES PLACE ON INTERACTION OF THE GLOBULAR HEADS OF C1Q WITH THE  
 CC FC REGIONS OF IGG OR IGM ANTIBODY PRESENT IN IMMUNE COMPLEXES.  
 CC -1- SUBUNIT: C1 IS A CALCIUM-DEPENDENT TRIMOLECULAR COMPLEX OF C1Q, R  
 CC AND S IN THE MOLAR RATION OF 1:2:2. C1Q SUBCOMPONENT IS COMPOSED  
 CC OF NINE SUBUNITS, SIX OF WHICH ARE DISULFIDE-LINKED DIMERS OF THE  
 CC A AND B CHAINS, AND THREE OF WHICH ARE DISULFIDE-LINKED DIMERS OF THE

CC THE C CHAIN. IN ADDITION TO THE MAJOR A:B AND C:C DIMER BANDS,  
 CC RAT, UNLIKE HUMAN C1Q, CONTAINED MINOR DIMER SPECIES.  
 CC -1- SIMILARITY: Contains 1 collagenous domain.  
 CC -1- SIMILARITY: Contains 1 C1Q domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ch/announce/>  
 CC or send an email to [license@ebi-sib.ch](mailto:license@ebi-sib.ch)).  
 CC -----  
 CC EMBL; X71127; CA50440.1; -.  
 CC PIR; S49158; S49158.  
 DR InterPro; IPR001073; C1q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00366; C1q; 1.  
 DR Pfam; PF01391; Collagen; 1.  
 DR PRINTS; PRO0007; COMPLEMENTC1Q.  
 DR SMART; SM00110; C1Q; 1.  
 DR PROSITE; PS01113; C1Q; 1.  
 KW Complement pathway; Plasma; Hydroxylation; Glycoprotein; Collagen;  
 KW Repeat; Signal.  
 FT SIGNAL 1 25  
 FT CHAIN 26 253  
 FT DOMAIN 29 112  
 FT DOMAIN 113 253  
 FT DISULFID 29 29  
 FT MOD\_RES 33 33  
 FT MOD\_RES 36 36  
 FT MOD\_RES 39 39  
 FT MOD\_RES 51 51  
 FT MOD\_RES 54 54  
 FT MOD\_RES 57 57  
 FT MOD\_RES 60 60  
 FT MOD\_RES 63 63  
 FT MOD\_RES 75 75  
 FT MOD\_RES 81 81  
 FT MOD\_RES 84 84  
 FT MOD\_RES 90 90  
 FT MOD\_RES 96 96  
 FT MOD\_RES 99 99  
 SQ SEQUENCE 253 AA; 26589 MW; 1CB40622571BFC9B CRC64;  
 Query Match 10.2%; Score 176.5; DB 1; Length 253;  
 Best Local Similarity 28.2%; Pred. No. 3e-07;  
 Matches 50; Conservative 40; Mismatches 66; Indels 21; Gaps 8;  
 QY 2 LPPLGLGLGPAACWALGP--TPGP-----SSSELR-----AFSAARTTPEGTSEMA 47  
 DB 80 IPGIPGVKVGK--PVGPKAPPPPPRGKGGSGDYKAKQKAFSLRLTVNSALARNQQA 137  
 QY 48 VTPEKXVYVINGDFDVAATGFCRVPAGYFFSTAGKAPKSLSVLVRNRDEVO-ALAF 106  
 DB 138 IREKEVITVNVNDYBPSGKFTCKPGGLYFTYHASSRGLCNVIVAGGRDRDMQKVLTF 197  
 QY 107 DEQRPRGARRAASQSMLOLDYGDITWLRHGAHPHALGAPAT--FSGVLVYVADDD 161  
 DB 198 CDYAAQ-NTFOVTTGGVVLKLEGGENVFLQATD-KNSLLGVEGANSIFGTGLPDDMD 252  
 RESULT 15  
 ID COLE IEPMMA STANDARD; PRT; 419 AA.  
 AC P98085; Q91080;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-1998 (Rel. 36, Last annotation update)  
 DE Inner ear-specific collagen precursor (Saccular collagen).  
 OS Lepomis macrochirus (Bluegill).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygi; Percomorpha; Perciformes; Percoidae;  
 OC Centrarchidae; Lepomis.  
 OX NCBI\_TaxID=13106;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95167486; PubMed=7863311;  
 RA Davis J.G., Oberholzer J.C., Burns F.R., Greene M.I.,  
 RT "Molecular cloning and characterization of an inner ear-specific  
 structural protein."  
 RL Science 267:1031-1034 (1995).  
 RN [2]  
 RP CONCEPTUAL TRANSLATION.  
 RA Gibson T.;  
 RL Submitted (MAR-1995) to the SWISS-PROT data bank.  
 CC -!- FUNCTION: FORMS A MICROSTRUCTURAL MATRIX WITHIN THE OTOLITHIC  
 CC MEMBRANE (PROBABLY).  
 CC -!- TISSUE SPECIFICITY: SPECIALIZED SECRETORY SUPPORTING CELLS AT THE  
 CC OUTER PERIMETER OF THE SACCULAR EPITHELIUM.  
 CC -!- SIMILARITY: contains 1 C1Q domain.  
 CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT HAD TO BE  
 CC INTRODUCED FOR POSITIONS 391-419 SO AS TO MAXIMIZE THE SIMILARITY  
 CC WITH OTHER SHORT-CHAIN COLLAGENS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC  
 DR EMBL; U17431; AA69978.1; ALT\_FRAME.  
 DR InterPro; IPR001073; C1Q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1q; 1.  
 DR PRINTS; PRO0007; COLLAGEN; 3.  
 DR PRODOM; PD000007; C1q\_helix; 2.  
 DR SMART; SM00110; C1Q; 1.  
 DR PROSITE; PS01113; C1Q; 1.  
 KW Extracellular matrix; Repeat; Collagen; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 419  
 FT DOMAIN 20 57  
 FT DOMAIN 58 274  
 FT DOMAIN 275 419  
 FT DOMAIN 272 419  
 FT CARBOHYD 37 37  
 FT CARBOHYD 320 320  
 FT CARBOHYD 320 320  
 SQ SEQUENCE 419 AA; 43634 MW; 570CDB9675FC0F39 CRC64;  
 Query Match 10.1%; Score 173.5; DB 1; Length 419;  
 Best Local Similarity 32.7%; Pred. No. 9, 4e-07;  
 Matches 56; Conservative 21; Mismatches 59; Indels 35; Gaps 9;  
 QY 5 LGLGPPAACWALGPTPGSS--ELRSAFSA---ARTTPEGTSEMAVTFDKYVING 58  
 DB 258 LKGVGGR-----GPKGPGESVEQIRSAFSGLFPSRSFP--PSSLVKEFKVFNGE 309  
 QY 59 GPDFVATGQRCRVPGAYFFSF--TAGKAPHKSL-----SVMLVRNDEVQALAFDEQRRP 112  
 DB 310 GMDPTLNKFNVTYPGVYLFYSYHITVRNRPVRAALVNGVTRKLRTRDSLYGODIDQ---- 365  
 QY 113 GARBAQSAMQLQDYGTVMRL---HGAPHYALGAPATGSGLYVAD 159  
 DB 366 -----ASNIALHLHTDGDQWLETLRDWNGX--YSSSEDDSTFGFLYFD 409



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:10 ; Search time 45.3114 Seconds  
(without alignments)  
1873.686 Million cell updates/sec

Title: US-10-085-167-2

Sequence: 1 MPELLGLGPAACWALGPT.....LYTPDLAPAAPPGALGASSELL 329

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP archaea: \*  
2: SP bacteria: \*  
3: SP fungi: \*  
4: SP human: \*  
5: SP invertebrate: \*  
6: SP mammal: \*  
7: SP mhc: \*  
8: SP organelle: \*  
9: SP phage: \*  
10: SP plant: \*  
11: SP rodent: \*  
12: SP virus: \*  
13: SP vertebrate: \*  
14: SP unclassified: \*  
15: SP virus: \*  
16: SP bacteriophage: \*  
17: SP archaea: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1718	99.5	329	4	Q81V25
2	1587	91.9	326	11	Q8R066
3	1003.5	58.1	205	11	Q9D0M2
4	997.5	57.8	205	11	Q9DCB6
5	230	13.3	158	4	Q9H667
6	230	13.3	1077	4	Q8TE71
7	226	13.1	158	11	Q8K110
8	220.5	12.8	182	11	Q8R1P2
9	220.5	12.8	281	11	Q9QXP7
10	213	12.3	243	4	Q8N6P2
11	207	12.0	294	11	Q9D8U4
12	205	11.9	243	11	Q8R002
13	202	11.7	243	11	Q8K479
14	199	11.5	243	6	Q95JD7
15	196	11.4	247	11	Q8BRW2
16	189	11.0	240	6	Q95M04

17	188	10.9	244	11	Q8K3R4	Q8K3R4 rattus norv
18	186	10.8	287	11	Q8CFR0	Q8CFR0 mus musculu
19	185.5	10.7	312	11	Q8CHX9	Q8CHX9 mus musculu
20	180.5	10.5	264	11	Q8BKRO	Q8BKRO mus musculu
21	179.5	10.4	120	11	Q8R1Z2	Q8R1Z2 mus musculu
22	176.5	10.2	289	11	Q8BVD7	Q8BVD7 mus musculu
23	170	9.8	194	6	Q95J95	Q95J95 canis fami1
24	168	9.7	245	11	Q9DCM6	Q9DCM6 mus musculu
25	167.5	9.7	744	11	Q921S8	Q921S8 mus musculu
26	167.5	9.7	744	11	Q8BGL6	Q8BGL6 mus musculu
27	166	9.6	675	6	Q9N178	Q9N178 sus scrofa
28	164	9.5	295	11	Q921K4	Q921K4 rattus norv
29	163	9.4	224	4	Q8ITK8	Q8ITK8 homo sapien
30	163	9.4	224	11	Q8BGU2	Q8BGU2 mus musculu
31	162.5	9.4	246	11	Q9ESJ0	Q9ESJ0 mus musculu
32	161	9.3	173	6	Q62789	Q62789 sus scrofa
33	155	9.0	196	11	Q920N0	Q920N0 tamias sibi
34	152	8.8	197	11	Q9JHGO	Q9JHGO mus musculu
35	150	8.7	333	4	Q8ITU4	Q8ITU4 homo sapien
36	145.5	8.4	705	4	Q8TEJ5	Q8TEJ5 homo sapien
37	141.5	8.2	1017	11	Q99K41	Q99K41 mus musculu
38	140	8.1	198	11	Q8BMP0	Q8BMP0 mus musculu
39	139	8.1	198	11	Q8BME9	Q8BME9 mus musculu
40	137.5	8.0	195	11	Q8BZS3	Q8BZS3 mus musculu
41	134.5	7.8	213	5	P83425	P83425 mytilus edu
42	132.5	7.7	347	4	Q961H6	Q961H6 homo sapien
43	131.5	7.6	185	5	Q9GQV4	Q9GQV4 strongyloce
44	131.5	7.6	583	4	Q96G58	Q96G58 homo sapien
45	131.5	7.6	992	4	Q9UC76	Q9UC76 homo sapien

## ALIGNMENTS

## RESULT 1

ID Q81V25 PRELIMINARY; PRT; 329 AA.

AC Q81V25;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Similar to C1q and tumor necrosis factor related protein 4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Strausberg R.;  
DR Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC035628; ANH35628.1; -  
SQ SEQUENCE 329 AA; 35256 MW; 16064DA8182A6732 CRC64;

Query Match 99.5%; Score 1718; DB 4; Length 329;

Best Local Similarity 99.7%; Pred. No. 2.2e-134;

Matches 328; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	MPELLGLGPAACWALGPTPPGSSSELSASASARTTPLEGTSEMAVFDKYVNIIGD	60
DB	1	MPELLGLGPAACWALGPTPPGSSSELSASASARTTLEGTSEMAVFDKYVNIIGD	60
QY	61	PVATGQFRCRVGAVFFSTAGKAPHKSLVWLNREDEVALAFDEQRPGARRASQ	120
DB	61	PDVATGQFRCRVGAVFFSTAGKAPHKSLVWLNREDEVALAFDEQRPGARRASQ	120
QY	121	SAMTOLDYDVTWLRHGAHYALGAPGATFSGYLVYADADADAPRGAPPEPSAFS	180
DB	121	SAMTOLDYDVTWLRHGAHYALGAPGATFSGYLVYADADADAPRGAPPEPSAFS	180
QY	181	AARTSLVSDAGPGRHQPPLAFDTEFVNIIGDFAAGVFRCLPGAYFFSTIGKLP	240
DB	181	AARTSLVSDAGPGRHQPPLAFDTEFVNIIGDFAAGVFRCLPGAYFFSTIGKLP	240

```

QY 241 KTLSTKLMKNRDEVOAMITVDGASRRRMOSSQVMLALRRGDVWMLSHDHGCGAYSNH 300
DB 241 KTLSTKLMKNRDEVOAMITVDGASRRRMOSSQVMLALRRGDVWMLSHDHGCGAYSNH 300
QY 301 GKTYTFSGFLVYPDLAPAPPGAGSEL 329
DB 301 GKTYTFSGFLVYPDLAPAPPGAGSEL 329
RESULT 2
QY 09R066 PRELIMINARY; PRT; 326 AA.
AC 09R066:
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Similar to C1q and tumor necrosis factor related protein 4.
GN 0710001E10R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 1
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027315; AAH27315.1; -.
DR MGI; MGI:1914695; 0710001E10R1K.
DR InterPro; IPR001073; C1q.
DR Pfam; PF00386; C1q; 2.
DR SMART; SM00110; C1q; 2.
DR PROSITE; PS01113; C1q; 2.
SQ SEQUENCE 326 AA; 35057 MW; 7233917287B1051A CRC64;
Query Match 91.9%; Score 1587; DB 11; Length 326;
Best Local Similarity 94.2%; Pred. No. 1.6e-123;
Matches 309; Conservative 2; Mismatches 15; Indels 2; Gaps 2;
QY 1 MLPLILGLGPAACALPTGPGSSSELRSAFSAARTTPELGTSMAATPFKYYVNIIGD 60
DB 1 MLPLILGLGPAACALGPA--GPGSSSELRSAFSAARTTPELGTSMAATPFKYYVNIIGD 59
QY 61 FDVATGPRCRVPGAYFSPFTAGKAPHSLSVMLVRNDEVQALAFDQRRPGARRASQ 120
DB 61 FDVATGPRCRVPGAYFSPFTAGKAPHSLSVMLVRNDEVQALAFDQRRPGARRASQ 119
QY 121 SAMQLDYGDTVWLRHGAHYALGAPGATFSGYLTVADADADAPARCPAPPEPRSAFS 180
DB 121 SAMQLDYGDTVWLRHGAHYALGAPGATFSGYLTVADADADAPARCPAPPEPRSAFS 178
QY 181 AARTSLVGSNDGPPRRHOPLAFTDFPNVIGDPFAAAGVPRCRPLGAYFSTLGLKLP 240
DB 181 AARTSLVGSNDGPPRRHOPLAFTDFPNVIGDPFAAAGVPRCRPLGAYFSTLGLKLP 238
QY 179 AARTSLVGSNDGPPRRHOPLAFTDFPNVIGDPFAAAGVPRCRPLGAYFSTLGLKLP 238
DB 179 AARTSLVGSNDGPPRRHOPLAFTDFPNVIGDPFAAAGVPRCRPLGAYFSTLGLKLP 238
QY 241 KTLSTKLMKNRDEVOAMITVDGASRRRMOSSQVMLALRRGDVWMLSHDHGCGAYSNH 300
DB 241 KTLSTKLMKNRDEVOAMITVDGASRRRMOSSQVMLALRRGDVWMLSHDHGCGAYSNH 298
QY 301 GKTYTFSGFLVYPDLAPAPPGAGSEL 328
DB 301 GKTYTFSGFLVYPDLAPAPPGAGSEL 326
RESULT 3
QY 09D0W2 PRELIMINARY; PRT; 205 AA.
AC 09D0W2:
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE 0710001E10R1K protein.
GN 0710001E10R1K.

```

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 1
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027315; AAH27315.1; -.
DR MGI; MGI:1914695; 0710001E10R1K.
DR InterPro; IPR001073; C1q.
DR Pfam; PF00386; C1q; 2.
DR SMART; SM00110; C1q; 2.
DR PROSITE; PS01113; C1q; 2.
SQ SEQUENCE 326 AA; 35057 MW; 7233917287B1051A CRC64;
Query Match 91.9%; Score 1587; DB 11; Length 326;
Best Local Similarity 94.2%; Pred. No. 1.6e-123;
Matches 309; Conservative 2; Mismatches 15; Indels 2; Gaps 2;
QY 1 MLPLILGLGPAACALPTGPGSSSELRSAFSAARTTPELGTSMAATPFKYYVNIIGD 60
DB 1 MLPLILGLGPAACALGPA--GPGSSSELRSAFSAARTTPELGTSMAATPFKYYVNIIGD 59
QY 61 FDVATGPRCRVPGAYFSPFTAGKAPHSLSVMLVRNDEVQALAFDQRRPGARRASQ 120
DB 61 FDVATGPRCRVPGAYFSPFTAGKAPHSLSVMLVRNDEVQALAFDQRRPGARRASQ 119
QY 121 SAMQLDYGDTVWLRHGAHYALGAPGATFSGYLTVADADADAPARCPAPPEPRSAFS 180
DB 121 SAMQLDYGDTVWLRHGAHYALGAPGATFSGYLTVADADADAPARCPAPPEPRSAFS 178
QY 181 AARTSLVGSNDGPPRRHOPLAFTDFPNVIGDPFAAAGVPRCRPLGAYFSTLGLKLP 240
DB 181 AARTSLVGSNDGPPRRHOPLAFTDFPNVIGDPFAAAGVPRCRPLGAYFSTLGLKLP 238
QY 179 AARTSLVGSNDGPPRRHOPLAFTDFPNVIGDPFAAAGVPRCRPLGAYFSTLGLKLP 238
DB 179 AARTSLVGSNDGPPRRHOPLAFTDFPNVIGDPFAAAGVPRCRPLGAYFSTLGLKLP 238
QY 241 KTLSTKLMKNRDEVOAMITVDGASRRRMOSSQVMLALRRGDVWMLSHDHGCGAYSNH 300
DB 241 KTLSTKLMKNRDEVOAMITVDGASRRRMOSSQVMLALRRGDVWMLSHDHGCGAYSNH 298
QY 301 GKTYTFSGFLVYPDLAPAPPGAGSEL 328
DB 301 GKTYTFSGFLVYPDLAPAPPGAGSEL 326
RESULT 3
QY 09D0W2 PRELIMINARY; PRT; 205 AA.
AC 09D0W2:
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE 0710001E10R1K protein.
GN 0710001E10R1K.

```

RC STRAIN=C57BL/6J; TISSUE=Brain;  
 RX MEDLINE=21085660; Pubmed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 Arakawa T., Hara A., Nishikishi Y., Konno H., Adachi J., Fukuda S.,  
 Aizawa K., Izawa M., Fushi K., Kiyosawa H., Kondo S., Yamataka I.,  
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochita H.,  
 Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 Schirml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Mashio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barah G.,  
 Blake J., Boffelli D., Bojunga N., Carinci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,  
 Guetincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombaeets P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,  
 Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohenki S.,  
 RA Hayaishiaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL: AK002948; BAB22473.1;  
 DR MGD: MGI:1914695; 0710001E10Rik.  
 DR InterPro: IPR001073; Clq.  
 DR Pfam: PF00386; Clq; 1.  
 DR PRINTS: PR00007; COMPLEMENTC1Q.  
 DR SMART: SM00110; Clq; 1.  
 DR PROSITE: PS01113; Clq; 1.  
 SQ SEQUENCE 205 AA; 22215 MW; 56AD37793C437300 CRC64;

Query Match 57.8%; Score 997.5; DB 11; Length 205;  
 Best Local Similarity 93.2%; Pred. No. 6, 4e-75;  
 Matches 192; Conservative 1; Mismatches 12; Indels 1; Gaps 1;

QY 123 MQLDYGDTWMLHGAHPHALGAPGATSGYLVADADADARGPAPPRPSAFSA 182  
 DB 1 MQLDYGDTWMLHGAHPHALGAPGATSGYLVADADADARGPAPPRPSAFSA 59  
 QY 183 RTRSLVGSAGPGRPHOPLAFDTEFNIGDPPDAAGVFCRLPGAYFSFTGLKLPRT 242  
 DB 60 RTRSLVGSAGPGRPHOPLAFDTEFNIGDPPDAAGVFCRLPGAYFSFTGLKLPRT 119  
 QY 243 LSVKLMKRNDEVQAMTYDDGASRRRMOQSVMALALRGDAVWLISHDHGCGAYSNHGK 302  
 DB 120 LSVKLMKRNDEVQAMTYDDGASRRRMOQSVMALALRGDAVWLISHDHGCGAYSNHGK 179  
 QY 303 YITFSGELVYPDLAPAPGAGSEL 328  
 DB 180 YITFSGELVYPDLAPAPGAGSEL 205

RESULT 5  
 Q9H667 PRELIMINARY; PRT; 158 AA.  
 AC Q9H667;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Hypothetical protein FLJ22569.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCB1 TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Small intestine;  
 RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,  
 RA Suzuki Y., Oobayashi M., Nishi T., Shibahara T., Tanaka T.,  
 RA Nakamura Y., Ieogai T., Sugano S.;  
 RT "NEO human cDNA sequencing project."  
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Uterus;  
 RA Strausberg R.;  
 RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK026222; BAB15398.1;  
 DR EMBL: BC007520; AA07520.1;  
 DR InterPro: IPR001073; Clq.  
 DR Pfam: PF00386; Clq; 1.  
 DR PRINTS: PR00007; COMPLEMENTC1Q.  
 DR SMART: SM00110; Clq; 1.  
 DR PROSITE: PS01113; Clq; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 158 AA; 17625 MW; 47DB10EDD6DC9760 CRC64;

Query Match 13.3%; Score 230; DB 4; Length 158;  
 Best Local Similarity 38.4%; Pred. No. 2e-11;  
 Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;  
 QY 163 DAPARP-----PAPPRPSAFSAARSLVSGAGGPRHOPPLAFTEFNIGG 212  
 DB 5 DVPVNPATILPVHYPLPQGMVAFSAAR-----SNLAGTLDQPIVFDLLNNGE 59  
 QY 213 DFDAAAGVRCRLPGAYFSFTGLKLP-RKTLVSKLMKRNDEVQAMTYDDGASRRRMOQS 271  
 DB 60 TFDLQGRNCFVNGTYVIFFMKLAVNVPLVNLKMEVLSAYANDGAP-DHETAS 118  
 QY 272 QSVMLALRRGDAVWLISHDHGCGA-YSNHGRTYITFSGELVYPD 314  
 DB 119 NNAIIQLFGQDIWRLH----RGAIVGSSWXYSTFSGYLLVQD 158

RESULT 6  
 Q8TE71 PRELIMINARY; PRT; 1077 AA.  
 AC Q8TE71;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE BEGIL.  
 GN EGG1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCB1 TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Aerajina W., Miller J.L.;  
 RL Submitted (Jan-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY074490; AAL71549.1;  
 DR InterPro: IPR001073; Clq.  
 DR Pfam: PF00386; Clq; 1.  
 DR SMART: SM00110; Clq; 1.  
 DR PROSITE: PS01113; Clq; 1.  
 SQ SEQUENCE 1077 AA; 120974 MW; 2B88BF3C47D032D6 CRC64;

Query Match 13.3%; Score 230; DB 4; Length 1077;  
 Best Local Similarity 38.4%; Pred. No. 2, 4e-10;  
 Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;  
 QY 163 DAPARP-----PAPPRPSAFSAARSLVSGAGGPRHOPPLAFTEFNIGG 212  
 DB 924 DVPVNPATILPVHYPLPQGMVAFSAAR-----SNLAGTLDQPIVFDLLNNGE 978  
 QY 213 DFDAAAGVRCRLPGAYFSFTGLKLP-RKTLVSKLMKRNDEVQAMTYDDGASRRRMOQS 271  
 DB 979 TFDLQGRNCFVNGTYVIFFMKLAVNVPLVNLKMEVLSAYANDGAP-DHETAS 1037  
 QY 272 QSVMLALRRGDAVWLISHDHGCGA-YSNHGRTYITFSGELVYPD 314  
 DB 1038 NNAIIQLFGQDIWRLH----RGAIVGSSWXYSTFSGYLLVQD 1077

RESULT 7

```

08K110
ID 08K110 PRELIMINARY; PRT; 158 AA.
AC 08K110:
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Similar to hypothetical protein FLJ22569.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA Straubeberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027523; AAH27523.1; -.
DR InterPro; IPR001073; Clq.
DR Pfam; PF00386; Clq; 1.
DR PRINTS; PRO0007; COMPLEMENTC1Q.
DR SMART; SM00110; Clq; 1.
DR PROSITE; PS01113; Clq; 1.
KW Hypothetical protein.
SQ SEQUENCE 158 AA; 17533 MW; 86E9321C99225FCB CRC64;

Query Match 13.1%; Score 226; DB 11; Length 158;
Best Local Similarity 39.1%; Pred. No. 4.2e-11;
Matches 63; Conservative 18; Mismatches 68; Indels 12; Gaps 5;

Qy 156 VYADADADAPARGPAPRPSAARTSLVSGDAGPGRHOPFLAFDTFVNIGDGD 215
Db 8 VTSPAAALPVHIVPLPQOMRVAFAART-----SNLAPGTLDPPIVFDLLNMLGTFN 62

Qy 216 AAAGVFCRLPGAFVFTSLGKLP-RKTLSTYKLMKNDEVOAMTYDDGASRRRMOGSV 274
Db 63 LQLRFCPCVNGTVFIFHMLKLVNVPVLYNLMKNESVLVSAYANDGAP-DHETASNHA 121

Qy 275 MLALRGDAVWLSSHHDGYGA-YSNHGKTYTFSGFLVYPD 314
Db 122 VLQLQGDQIVLRH-----RRAIVGSSWKSTFGVLYLYPD 158

RESULT 8
08R1P2 PRELIMINARY; PRT; 182 AA.
ID 08R1P2:
AC 08R1P2:
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Similar to RIKEN CDNA 1600017K21 gene (Fragment).
GN 1600017K21RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Straubeberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC023468; AAH23468.1; -.
DR MGD; MGI:1919254; 1600017K21RIK.
DR InterPro; IPR001073; Clq.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF00386; Clq; 1.
DR SMART; SM00110; Clq; 1.
FT NON TER
SQ SEQUENCE 182 AA; 20863 MW; 559C73DE9517882F CRC64;

Query Match 12.8%; Score 220.5; DB 11; Length 182;
Best Local Similarity 39.2%; Pred. No. 1.4e-10;
Matches 56; Conservative 20; Mismatches 58; Indels 9; Gaps 4;

```

```

Qy 177 SAFSAARTSLVSGDAGPGRHOPFLAFDTFVNIGDGDPAAGVFCRLPGAFVFTSLG 236
Db 47 AAFVSGRRKALNSND-----YFQPVVFDTEFVNLGHFMNFGKFCYCYGVYFSLNHA 101

Qy 237 KLPKRLTSVKLMKNREVOAMTYDDGASRRRMOGSVWLALRGDAVW--LLSHHDGY 294
Db 102 TNNQKTYVLIHIMNNEEV-VILTAQ-VSDRSITWOSLMMELREEDBVVRLFKGRENA 159

Qy 295 GAYSNHGKTYTFSGFLVYPDLP 317
Db 160 IFSDEFDTYITTFSGVLVKNPASEP 182

RESULT 9
09QXP7 PRELIMINARY; PRT; 281 AA.
ID 09QXP7:
AC 09QXP7:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative secreted protein ZSIG37 (1600017K21RIK protein).
GN ZSIG37 OR 1600017K21RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Sheppard P., Delisher T., Grant F., Chen L., Haldeman B., McKnight G.,
RA Whitmore T., O'Hare P.; secreted protein."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Placenta;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Atakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Guericich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilting L.,
RA Wyrshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohatsu S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690 (2001).
DR EMBL; AF192499; AAF06664.1; -.
DR EMBL; AK005484; BAB24070.1; -.
DR MGD; MGI:1919254; 1600017K21RIK.
DR InterPro; IPR001073; Clq.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF00386; Clq; 1.
DR Pfam; PF01391; Collagen; 1.
DR PRINTS; PRO0007; COMPLEMENTC1Q.
DR SMART; SM00110; Clq; 1.
SQ SEQUENCE 281 AA; 32009 MW; C9816216DB6419E2 CRC64;

Query Match 12.8%; Score 220.5; DB 11; Length 281;
Best Local Similarity 39.2%; Pred. No. 2.5e-10;
Matches 56; Conservative 20; Mismatches 58; Indels 9; Gaps 4;

Qy 177 SAFSAARTSLVSGDAGPGRHOPFLAFDTFVNIGDGDPAAGVFCRLPGAFVFTSLG 236

```

Db 146 AAFSVGRKKALHSND-----YFGPVVPTDFVNLKXHPNMTGFKCYCVPICIVFFSLNVH 200  
 QY 237 KLPKKTSLVKLMKRDVEQAMITYDGASRRREMOSQSVMLALRGDAVM--LLSHDHGY 294  
 Db 201 TWMOKETYLHIHMKREEV-VLLYAQ-VSDRSIMOSQSLMELRDEDEVMVWLLFGGERENA 258  
 QY 295 GAVSNHGKITYFSGFLVYPPDLAP 317  
 Db 259 IFSDEFTYITFSGFLVYKPASEP 281

RESULT 10  
 Q8N6P2 PRELIMINARY; PRT; 243 AA.  
 AC Q8N6P2;  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE C1q and tumor necrosis factor related protein 5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Strausberg R.;  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC029485; AAH29485.1; -  
 DR InterPro; IPR001073; C1q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1q; 1.  
 DR Pfam; PF01391; Collagen; 1.  
 DR PRINTS; PR00007; COMPLEMENTC1Q.  
 DR SMART; SM00110; C1Q; 1.  
 SQ SEQUENCE 243 AA; 25326 MW; 6D9306A0EB21B44A CRC64;

Query Match 12.3%; Score 213; DB 4; Length 243;  
 Best Local Similarity 34.9%; Pred. No. 8.8e-10;  
 Matches 66; Conservative 18; Mismatches 83; Indels 22; Gaps 6;

QY 138 GAPHYALGARGATSSG---YLVAADAD---ADAPAGCPAPP-----EPSSAFSAKTR 165  
 Db 57 GAP-----GAPGEGEGGPGPLGPRGPGGPRGEGGPAFGPGAGECVPSPRSASAKRSE 112  
 QY 186 SLVGSAGAPGPRHQPLAFDTFVNIGDPDAAGVFCRLPGAYFFSFTLGLPKRTLSV 245  
 Db 113 SRV-----PPSDAPLPFDVNLVNEGCHYDAVTKFTCYVPGYVFA-VNATVYRASLOF 166  
 QY 246 KLMKRDVEQAMITYDGASRRREMOSQSVMLALRGDAVWLLSHDHGAYSNHGKTYT 305  
 Db 167 DLVKNGESIASFQFGGWPKPASLSGAMVRLPEPDQVWVQVGVGDYIGIYASIKTDSI 226  
 QY 306 FSGFLVYPP 314  
 Db 227 FSGFLVYSD 235

RESULT 11  
 Q8D8U4 PRELIMINARY; PRT; 294 AA.  
 AC Q8D8U4;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE 1810033K05Rik protein (RIKEN cDNA 1810033K05 gene).  
 GN 1810033K05Rik.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=CS7BL/6J; TISSUE=Pancreas;  
 MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,  
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuoka H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Giesl C., King B., Kochiwa H.,  
 RA Kiehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carrini P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seyer T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
 RA Hayashizaki Y.,  
 RA "Functional annotation of a full-length mouse cDNA collection.";  
 RT Nature 409:685-690(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK007683; BAB25187.1; -  
 DR EMBL; BC030324; AAH30324.1; -  
 DR MGI; MGI:1916433; 1810033K05Rik.  
 DR InterPro; IPR001073; C1q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1q; 1.  
 DR Pfam; PF01391; Collagen; 2.  
 DR PRINTS; PR00007; COMPLEMENTC1Q.  
 DR PROSITE; PS01113; C1Q; 1.  
 SQ SEQUENCE 294 AA; 30865 MW; 6D3905AE7C19BE6A CRC64;

Query Match 12.0%; Score 207; DB 11; Length 294;  
 Best Local Similarity 33.7%; Pred. No. 3.6e-09;  
 Matches 55; Conservative 31; Mismatches 47; Indels 30; Gaps 8;

QY 7 GLIGPACNALGTPPGSGSELASAPSA--RTTPLEGTSSEMAVTDPKYVNTGGPDVA 64  
 Db 145 GLPGPSC-----GSSRAKSAFSAVATKSYPRE--RLPIKFKILMNEGSHYNAS 192  
 QY 65 TGGFRGVGAYPFSTAGAPKKSLSVMLVRRRDEQVALAFEGRRPARGAASQSAWL 124  
 Db 193 SGKFTVSVPEIYFTYDITLA-NKHAIIGLVHN-GQIRITFD--ANTGNHVAAGSTIL 248  
 QY 125 QLDYGDVWLRHGA-----PHYALGAPGATFSGYLVYAD 159  
 Db 249 ALKEGDGVWLRIFYSQNGLFFYDPVWT---DSLFGFLIYAD 267

RESULT 12  
 Q8R002 PRELIMINARY; PRT; 243 AA.  
 AC Q8R002;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Similar to DKFP2586B0621 protein (Hypothetical 25.4 kDa protein).  
 GN C1QTF5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.

RA Strausberg R.;  
 RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: BC023068; AAH23068.1; -  
 DR EMBL: BC025174; AAH25174.1; -  
 DR MGD: MGI:2385958; Clqnf5.  
 DR InterPro: IPR001073; Clq.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF00366; Clq; 1.  
 DR Pfam: PF01391; Collagen; 1.  
 DR SMART: SM00110; Clq; 1.  
 DR Hypothetical protein.  
 KW  
 SQ SEQUENCE 243 AA; 25420 MW; 498129CD051DB97B CRC64;

Query Match 11.9%; Score 205; DB 11; Length 243;  
 Best Local Similarity 32.8%; Pred. No. 4,1e-09;  
 Matches 64; Conservative 16; Mismatches 81; Indels 34; Gaps 6;

QY 138 GAPHYALGAPGATGSGYLIVADADADAPA-RGPPAP-----DEPRSAF 179  
 DB 57 GAP-----GAPGKGEK-----GRPGLPGRPGEPGRGEGAPMGALGPAGECSVPPRSAP 106  
 QY 180 SAATRSIVGSDAGPGRHOPLAFTDFEVNIGDFDAAAGVFRCLGAYFFSTLGLKP 239  
 DB 107 SAKRSERV-----PPPADTPLPFDRVLNMQGHFDTTGKFTQVPGVYFA-VHATVY 160  
 QY 240 RKTLSVKLMKNRDEVQAMTYDDGASRRREMOSQSVMLALRGDAVWLHSHDHGCGAYSN 299  
 DB 161 RASIQFDLVKNGOSIASFFQYFGGWPKPASLSGAMVRLPEPDQVWVQGVGDIYGIYAS 220  
 QY 300 HGKYTFSGFLVYPD 314  
 DB 221 IKTDSTFGFLVYSD 235

RESULT 13  
 Q8K479 PRELIMINARY; PRT; 243 AA.

AC O8K479;  
 DT 01-OCT-2002 (TREMBLrel. 22, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Complement-c1q tumor necrosis factor-related protein.  
 GN ClqTNF5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J;  
 RX MEDLINE=22135657; PubMed=12140190;  
 RA Kaneva S., Hawes N.L., Chang B., Heckenlively J.R., Naggert J.K.,  
 Nishina P.M.;  
 RT "Mrip, a gene encoding a frizzled related protein, is mutated in the  
 mouse retinal degeneration 6.";  
 RL Hum. Mol. Genet. 11:1879-1886(2002).  
 DR EMBL: AF469650; AAM69217.1; -  
 DR MGD: MGI:2385958; Clqtnf5.  
 DR InterPro: IPR001073; Clq.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF00366; Clq; 1.  
 DR Pfam: PF01391; Collagen; 2.  
 DR PRINTS: PR000007; COMPLEMENTC1Q.  
 DR SMART: SM00110; Clq; 1.  
 SQ SEQUENCE 243 AA; 25436 MW; 9FAD5604349791D9 CRC64;

Query Match 11.7%; Score 202; DB 11; Length 243;  
 Best Local Similarity 32.3%; Pred. No. 7,2e-09;  
 Matches 63; Conservative 17; Mismatches 81; Indels 34; Gaps 6;

QY 138 GAPHYALGAPGATGSGYLIVADADADAPA-RGPPAP-----DEPRSAF 179  
 DB 57 GAP-----GAPGKGEK-----GRPGLPGRPGEPGRGEGAPMGALGPAGECSVPPRSAP 106

QY 180 SAATRSIVGSDAGPGRHOPLAFTDFEVNIGDFDAAAGVFRCLGAYFFSTLGLKP 239  
 DB 107 SAKRSERV-----PPPADTPLPFDRVLNMQGHFDTTGKFTQVPGVYFA-VHATVY 160  
 QY 240 RKTLSVKLMKNRDEVQAMTYDDGASRRREMOSQSVMLALRGDAVWLHSHDHGCGAYSN 299  
 DB 161 RASIQFDLVKNGOSIASFFQYFGGWPKPASLSGAMVRLPEPDQVWVQGVGDIYGIYAS 220  
 QY 300 HGKYTFSGFLVYPD 314  
 DB 221 IKTDSTFGFLVYSD 235

RESULT 14

Q95JD7 PRELIMINARY; PRT; 243 AA.  
 AC Q95JD7;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Adiponectin.  
 GN APM1.  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecoidea; Macaca.  
 CX NCBI\_TaxID=9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Adipose tissue;  
 RX MEDLINE=2123234; PubMed=11334417;  
 RA Horta K., Funahashi T., Bodkin N.L., Ortmeier H.K., Arlt Y.,  
 Hansen B.C., Matsuzawa Y.;  
 RT "Circulating concentrations of the adipocyte protein adiponectin are  
 decreased in parallel with reduced insulin sensitivity during the  
 progression to type 2 diabetes in rhesus monkeys.";  
 RL Diabetes 50:1126-1133(2001).  
 DR EMBL: AF404407; AK62202.1; -  
 DR InterPro: IPR001073; Clq.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF00366; Clq; 1.  
 DR Pfam: PF01391; Collagen; 1.  
 DR PRINTS: PR000007; COMPLEMENTC1Q.  
 DR ProDom: PD000007; Collagen; 1.  
 DR SMART: SM00110; Clq; 1.  
 DR PROSITE: PS01113; Clq; 1.  
 KW Collagen.  
 SQ SEQUENCE 243 AA; 26264 MW; 49A45DAF2B4613FD CRC64;

Query Match 11.5%; Score 199; DB 6; Length 243;  
 Best Local Similarity 29.7%; Pred. No. 1,3e-08;  
 Matches 54; Conservative 31; Mismatches 61; Indels 36; Gaps 8;

QY 7 GLIGP-----AACWALGP-----TEPGSSSELRSAPSAARTPLEGTEMAV 48  
 DB 71 GLIGPKDGTGETGTGTAEGPGRPGIQRKGEPEGAYVYRSAPSVGLEITYVT-VPNNPI 129  
 QY 49 TFDKVVYVNIIGDPEVALAGQPRCRVPAGVFFSFTGKXPH-----KSLSVMLVRRNDEVQA 103  
 DB 130 RFTKIFVYNOQNHVYGSFGKFCNIPGLYFAY-----HTVYMKDVKVSLEK-KDKXML 182  
 QY 104 LAPDEORRPGARRAASOSAMLOLDYGTVMRLHGAPH-----VALGAPGATFSGYLIVAD 159  
 DB 183 FTYQYQENNVQD-ASGSVLLHLEVQVWLVQVYGEGERNGLYADNNDSTFTGFLYHD 241

RESULT 15  
 Q8BRW2

ID Q8BRW2 PRELIMINARY; PRT; 247 AA.  
 AC Q8BRW2;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DE Adipocyte complement related protein of 30 kDa.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; Tissue=Aorta and vein;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK041214; BAC30866.1; -.  
 SQ SEQUENCE 247 AA; 26751 MW; 0D3FA64C789CAEF3 CRC64;

Query Match 11.4%; Score 196; DB 11; Length 247;  
 Best Local Similarity 31.9%; Pred. No. 2.3e-08;  
 Matches 58; Conservative 30; Mismatches 58; Indels 36; Gaps 11;

QY 7 GLIGP-----AACWALGP-----TPG-----PGSSE--LRSAFSAARTTPELGTSSEMAV 48  
 DB 75 GLIGPKGETGVGVMGTGAEGPRGPGTGRKGEPEGAAYVYRSASFVGLETRVTV-VPNPVI 133  
 QY 49 TFDKVVYVNI GCDPVDATGFRCEVPGAYFFSFTAGKAPH-----KSLSVMLVNRNDEVOA 103  
 DB 134 RTTKLIFYNQNHVDSGTGKFCYCNIPGLYYFSY-----HITVYMKDVKVSLEPK-KDKAVL 186  
 QY 104 IAFDEQRPPGARRAASQSAMLQLDYDGTWMLRLHG-APH---YALGAPGATPSGYLVYAD 159  
 DB 187 FTYDQYQEKNVQD-QASGSVTLHLIEVGDQYWLQVYGGDGHNGLYADNVNDSTFTGFLYHD 245  
 QY 160 AD 161  
 DB 246 TN 247

Search completed: January 12, 2004, 08:19:07  
 Job time : 49.3114 secs

**THIS PAGE BLANK (USPTO)**



PR 01-JUL-1999; 99US-0346502.  
XY

PA (ZYMO) ZYMOGENETICS INC.  
 XX Holloway JL, Lok S;  
 XX MPI: 2001-138140/14.  
 DR N-PSDB; AAF28672.  
 XX Novel secreted protein ZACRP4 polypeptides having tandem C1q globular  
 PT domains, useful for studying cell-cell communication and regulation of  
 PT cellular processes -  
 XX  
 PS Claim 1; Page 77-78; 82pp; English.  
 XX  
 CC The present sequence is human ZACRP4 protein. ZACRP4 protein has two  
 CC complement factor C1q domains. The ZACRP4 gene is located on human  
 CC chromosome 11q11. The ZACRP4 coding sequence and protein have a number of  
 CC uses described in the specification, including, modulation of energy  
 CC balance and cellular metabolic reactions in mammals. In addition, ZACRP4  
 CC protein is useful as an autocrine factor, particularly during  
 CC development, in mediating the processes of an organism, in regulating  
 CC cellular processes such as cell proliferation and/or differentiation,  
 CC cell survival and energy balance.  
 CC  
 XX Sequence 329 AA;  
 SQ  
 Query Match 100.0%; Score 1726; DB 22; Length 329;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-154;  
 Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MLPLLLGLGPAACWALGPTPGSSSELSAFAARTTPELGTSMAVTFDKVYVNIIGD 60  
 DB 1 MLPLLLGLGPAACWALGPTPGSSSELSAFAARTTPELGTSMAVTFDKVYVNIIGD 60  
 QY 61 FVATGQRCRCVPGAYFFSFTAGKAPKXSLVMLVRNDEVQALAFDEQRPGARRASQ 120  
 DB 61 FVATGQRCRCVPGAYFFSFTAGKAPKXSLVMLVRNDEVQALAFDEQRPGARRASQ 120  
 QY 121 SAMQLDVGDTVWRLHGAHYALGAPGATFSGYLVAADADADAPARGPAPPEPRSAFS 180  
 DB 121 SAMQLDVGDTVWRLHGAHYALGAPGATFSGYLVAADADADAPARGPAPPEPRSAFS 180  
 QY 121 SAMQLDVGDTVWRLHGAHYALGAPGATFSGYLVAADADADAPARGPAPPEPRSAFS 180  
 DB 121 SAMQLDVGDTVWRLHGAHYALGAPGATFSGYLVAADADADAPARGPAPPEPRSAFS 180  
 QY 181 AARTSLVGSNAGPGRHOPALFDTFVNIGDPPDAAGVFCRLPGAVFSTTGKLP 240  
 DB 181 AARTSLVGSNAGPGRHOPALFDTFVNIGDPPDAAGVFCRLPGAVFSTTGKLP 240  
 QY 181 AARTSLVGSNAGPGRHOPALFDTFVNIGDPPDAAGVFCRLPGAVFSTTGKLP 240  
 DB 181 AARTSLVGSNAGPGRHOPALFDTFVNIGDPPDAAGVFCRLPGAVFSTTGKLP 240  
 QY 241 KTLVVKLMKNRDEVQAMITVDGASRRRQSQSWMLALBRGAWMLSHDHGYSN 300  
 DB 241 KTLVVKLMKNRDEVQAMITVDGASRRRQSQSWMLALBRGAWMLSHDHGYSN 300  
 QY 301 GKYITFSGLVYPDLAPAPPGIGASELL 329  
 DB 301 GKYITFSGLVYPDLAPAPPGIGASELL 329  
 QY 301 GKYITFSGLVYPDLAPAPPGIGASELL 329  
 DB 301 GKYITFSGLVYPDLAPAPPGIGASELL 329  
 RESULT 2  
 ABG79643  
 ID ABG79643 standard; Protein; 329 AA.  
 XX  
 AC ABG79643;  
 XX  
 DT 15-NOV-2002 (first entry)  
 XX  
 DE Human novel secreted protein SECP19, Incyte ID No. 931619CD1.  
 XX  
 XX Human; SECP, secreted protein; cell proliferative disorder;  
 KM actinic keratosis; arteriosclerosis; bursitis; hepatitis; cancer;  
 KM autoimmune disorder; inflammatory disorder; AIDS; asthma; allergy;  
 KM acquired immunodeficiency syndrome; anaemia; atopic dermatitis;  
 KM cardiovascular disorder; congestive heart failure; vascular tumour;  
 KM ischaemic heart disease; myocardial infarction; epilepsy; stroke;  
 KM hyperaesthetic heart disease; neurological disorder; cerebral neoplasm;  
 KM Alzheimer's disease; developmental disorder; renal tubular acidosis;  
 KM Cushing's syndrome; Duchenne muscular dystrophy; hypothyroidism;

KM Becker muscular dystrophy.  
 XX  
 OS Homo sapiens.  
 XX WO200262841-A2.  
 XX  
 XX 15-AUG-2002.  
 PD  
 XX 28-JAN-2002; 2002WO-US02616.  
 PF  
 XX 02-FEB-2001; 2001US-266195P.  
 PR 08-FEB-2001; 2001US-267924P.  
 PR 09-FEB-2001; 2001US-267816P.  
 PR 09-FEB-2001; 2001US-268112P.  
 PR 26-FEB-2001; 2001US-271639P.  
 PR 07-SEP-2001; 2001US-317818P.  
 PR 21-DEC-2001; 2001US-343553P.  
 XX  
 PA (INCYTE) INCYTE GENOMICS INC.  
 XX  
 PI Tang TY, Yue H, Gandhi AR, Yao MG, Warren BA, Ding L, Duggan BW,  
 PI Xu Y, Yang J, Thangavelu K, Lal PG, Honchell CD, Walla NK, Lee S,  
 PI Lee EA, Richardson TW, Baughn MR, Elliott VS;  
 XX  
 DR MPI: 2002-657522/70.  
 DR N-PSDB; ABS64954.  
 XX  
 PT New human secreted proteins and nucleic acids useful in diagnosing,  
 PT treating and preventing cell proliferative, autoimmune/inflammatory,  
 PT cardiovascular, neurological, and developmental disorders -  
 PT  
 PT Claim 1; Page 140; 158pp; English.  
 XX  
 XX The invention relates to twenty four human secreted proteins  
 CC (SECP1-24), proteins 90% identical to them and active fragments of them.  
 CC Also included are nucleic acids encoding the SECP proteins, a recombinant  
 CC polynucleotide comprising a promoter sequence operably linked to the  
 CC nucleic acid, a cell transformed with the recombinant polynucleotide, an  
 CC transgenic organism comprising the recombinant polynucleotide, an  
 CC anti-SECP antibody, and screening for ant/agonists and modulators of  
 CC SECP function or expression. The SECP proteins and nucleic acids are  
 CC useful in the diagnosis, treatment and prevention of cell proliferative  
 CC (e.g. actinic keratosis, arteriosclerosis, bursitis, hepatitis or  
 CC cancer), autoimmune/inflammatory (e.g. AIDS (acquired immunodeficiency  
 CC syndrome), asthma, anaemia, allergies or atopic dermatitis),  
 CC cardiovascular (e.g. congestive heart failure, ischaemic heart disease,  
 CC myocardial infarction, hypertensive heart disease, or vascular tumours),  
 CC neurological (e.g. epilepsy, stroke, cerebral neoplasms, or Alzheimer's  
 CC disease), and developmental (e.g. renal tubular acidosis, Cushing's  
 CC syndrome, Duchenne and Becker muscular dystrophy, or hypothyroidism)  
 CC disorders. Many other diseases and disorders are listed in the  
 CC specification. These may also be used in assessing the effects of  
 CC exogenous compounds on the expression of nucleic acid and amino acid  
 CC sequences of the secreted proteins. The present sequence represents a  
 CC SECP protein of the invention.  
 CC  
 XX Sequence 329 AA;  
 SQ  
 Query Match 98.9%; Score 1707; DB 23; Length 329;  
 Best Local Similarity 99.4%; Pred. No. 7.6e-153;  
 Matches 327; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MLPLLLGLGPAACWALGPTPGSSSELSAFAARTTPELGTSMAVTFDKVYVNIIGD 60  
 DB 1 MLPLLLGLGPAACWALGPTPGSSSELSAFAARTTPELGTSMAVTFDKVYVNIIGD 60  
 QY 61 FVATGQRCRCVPGAYFFSFTAGKAPKXSLVMLVRNDEVQALAFDEQRPGARRASQ 120  
 DB 61 FVATGQRCRCVPGAYFFSFTAGKAPKXSLVMLVRNDEVQALAFDEQRPGARRASQ 120  
 QY 121 SAMQLDVGDTVWRLHGAHYALGAPGATFSGYLVAADADADAPARGPAPPEPRSAFS 180  
 DB 121 SAMQLDVGDTVWRLHGAHYALGAPGATFSGYLVAADADADAPARGPAPPEPRSAFS 180

```
QY 181 AARTSLVSGSDAGPGRHQPPLAFDTEFVNIGGDFDAAGVFRCKLPAGVFFSFTLGKLP 240
DB 181 AARTSLVSGSDAGPGRHQPPLAFDTEFVNIGGDFDAAGVFRCKLPAGVFFSFTLGKLP 240
QY 241 KTLSTVLMKMRDEVQAMITYDDGASRRREMOSQSVMLALRRGDAVWLISHDHGCGAYSNNH 300
DB 241 KTLSTVLMKMRDEVQAMITYDDGASRRREMOSQSVMLALRRGDAVWLISHDHGCGAYSNNH 300
QY 301 GKTYTFSGFLVYPDLAPAAPPGIGASELL 329
DB 301 GKTYTFSGFLVYPDLAPAAPPGIGASELL 329

RESULT 3
AAB61424
ID AAB61424 standard; protein; 329 AA.
XX
AC AAB61424;
XX
DT 04-APR-2001 (first entry)
XX
DE Monkey MANGO 245 protein.
XX
KW TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;
KW autoimmune; allergy; cardiovascular; brain; degenerative; placental;
KW pancreatic; skeletal; muscle.
XX
OS Catarhini sp..
XX
PN WO200100672-A1.
XX
PD 04-JAN-2001.
XX
PF 29-JUN-2000; 2000WO-US18184.
XX
PR 29-JUN-1999; 99US-0342687.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;
XX
WPI; 2001-050127/06.
XX
PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver
PT disorders (e.g. jaundice) -
XX
PS Claim 1; Fig 25; 262pp; English.
XX
CC The present invention relates to cDNAs encoding TANGO 244,
CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.
CC The nucleic acids, proteins and protein modulators are useful for
CC treating colonic disorders, inflammatory diseases, tumors,
CC renal disorders, liver disorders, lung disorders, autoimmune diseases,
CC allergic diseases, cardiovascular diseases, brain disorders,
CC degenerative diseases placental, pancreatic, skeletal and muscle
CC disorders.
XX
SQ Sequence 329 AA;
XX

Query Match 96.5%; Score 1666; DB 22; Length 329;
Best Local Similarity 96.7%; Pred. No. 5,7e-149;
Matches 318; Conservative 1; Mismatches 10; Indels 0; Gaps 0;
```

```
QY 121 SAMLQLDYSDYDTWMLRLHGAHYALGAPGATFSGYLVYADADADAPARGPAPPEPSAFA 180
DB 121 SAMLQLDYSDYDTWMLRLHGAHYALGAPGATFSGYLVYADADADAPARGPAPPEPSAFA 180
QY 181 AARTSLVSGSDAGPGRHQPPLAFDTEFVNIGGDFDAAGVFRCKLPAGVFFSFTLGKLP 240
DB 181 AARTSLVSGSDAGPGRHQPPLAFDTEFVNIGGDFDAAGVFRCKLPAGVFFSFTLGKLP 240
QY 241 KTLSTVLMKMRDEVQAMITYDDGASRRREMOSQSVMLALRRGDAVWLISHDHGCGAYSNNH 300
DB 241 KTLSTVLMKMRDEVQAMITYDDGASRRREMOSQSVMLALRRGDAVWLISHDHGCGAYSNNH 300
QY 301 GKTYTFSGFLVYPDLAPAAPPGIGASELL 329
DB 301 GKTYTFSGFLVYPDLAPAAPPGIGASELL 329

RESULT 4
AAB61423
ID AAB61423 standard; protein; 348 AA.
XX
AC AAB61423;
XX
DT 04-APR-2001 (first entry)
XX
DE Human MANGO 245 protein.
XX
KW TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;
KW autoimmune; allergy; cardiovascular; brain; degenerative; placental;
KW pancreatic; skeletal; muscle.
XX
OS Homo sapiens.
XX
PN WO200100672-A1.
XX
PD 04-JAN-2001.
XX
PF 29-JUN-2000; 2000WO-US18184.
XX
PR 29-JUN-1999; 99US-0342687.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;
XX
WPI; 2001-050127/06.
XX
PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver
PT disorders (e.g. jaundice) -
XX
PS Claim 1; Fig 23; 262pp; English.
XX
CC The present invention relates to cDNAs encoding TANGO 244,
CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.
CC The nucleic acids, proteins and protein modulators are useful for
CC treating colonic disorders, inflammatory diseases, tumors,
CC renal disorders, liver disorders, lung disorders, autoimmune diseases,
CC allergic diseases, cardiovascular diseases, brain disorders,
CC degenerative diseases placental, pancreatic, skeletal and muscle
CC disorders.
XX
SQ Sequence 348 AA;
XX

Query Match 90.8%; Score 1567; DB 22; Length 348;
Best Local Similarity 99.7%; Pred. No. 1.4e-139;
Matches 229; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 61 FDVATGQFRCPVPGAYFFSFTAGKAPHKSLSVMLVRNDEVQALAFDEQRPGARRASQ 120  
DB 61 FDVATGQFRCPVPGAYFFSFTAGKAPHKSLSVMLVRNDEVQALAFDEQRPGARRASQ 120  
QY 121 SAMLOLDYGTWVRLHGAAPHYALGAPGATFSGYLVVADADADAPARPPAPPPRRSAFS 180  
DB 121 SAMLOLDYGTWVRLHGAAPHYALGAPGATFSGYLVVADADADAPARPPAPPPRRSAFS 180  
QY 181 AARTSLVGSAGCPGPRHQPPLAFTEFVNIIGDGFDAAGVRCRLPGAYFFSFTLGLPR 240  
DB 181 AARTSLVGSAGCPGPRHQPPLAFTEFVNIIGDGFDAAGVRCRLPGAYFFSFTLGLPR 240  
QY 241 KTLISVKLMKNRDEVQAMITYDDGASRRRMOSSQSVMLALRRGDVAWLSSHHDGYGAYSNH 300  
DB 241 KTLISVKLMKNRDEVQAMITYDDGASRRRMOSSQSVMLALRRGDVAWLSSHHDGYGAYSNH 300  
RESULT 5  
ABG70385  
ID ABG70385 standard; Protein; 299 AA.  
AC ABG70385;  
XX  
DT 05-NOV-2002 (first entry)  
DE Adipocyte complement-related C1q Tumour Necrosis Factor-like protein #3.  
XX  
KW Human; NOX; NOX-associated disorder; cardiomyopathy; atherosclerosis;  
cell signal processing; metabolic pathway modulation; metabolic disorder;  
obesity; diabetes; infectious disease; neurodegenerative disorder; acne;  
Alzheimer's disease; Parkinson's disease; immune disorder; cancer;  
haematopoietic disorder; cirrhosis; pancreatitis; learning defect;  
memory defect; infertility; congenital heart defect; hair growth;  
pigmentation disorder; endocrine disorder; respiratory disease; health;  
gastro-intestinal disease; reproductive; neurological disease;  
bone marrow transplantation; endocrine disease; allergy; inflammation;  
neurological disorder; urinary system disorder; age-related disorder;  
neuropsychiatric disorder; EGF-related protein; SCUB2; TEN-M4;  
adipocyte complement-related C1q tumour necrosis factor; out at first;  
beta adrenergic receptor kinase; EphA6/ehk-2; glucose transporter;  
type Ia membrane suah1-containing domain; butyrophilin;  
type Ia membrane suah1 domain containing.  
XX  
OS Homo sapiens.  
XX  
PN WO200257453-A2.  
XX  
PD 25-JUL-2002.  
XX  
PF 19-DEC-2001; 2001WO-US50331.  
XX  
PR 19-DEC-2000; 2000US-265704P.  
XX  
PR 20-DEC-2000; 2000US-257114P.  
PR 02-MAY-2001; 2001US-288153P.  
PR 29-MAY-2001; 2001US-294075P.  
PR 24-JUL-2001; 2001US-307506P.  
PR 10-AUG-2001; 2001US-311590P.  
PR 10-AUG-2001; 2001US-311613P.  
PR 29-AUG-2001; 2001US-315617P.  
PR 14-SEP-2001; 2001US-322358P.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Gangoli EA, Patturajan M, Vernet CAM, Malvanekar UM, Kekuda R;  
PI Stone DJ, Anderson D, Shimkets RA, Burgess CE, Zernhuken BD, Liu X;  
PI Spytek KA, Casman SO, Boldog FL, Smithson G, Li L, Ji W;  
XX  
DR MPI; 2002-590744/63.  
DR N-PSDB; ABSS2097.  
XX  
PT Novel isolated NOX polypeptide useful for treating cardiomyopathy,  
atherosclerosis, metabolic disorders, diabetes, obesity, infectious

PT disease, anorexia, neurodegenerative disorders, Alzheimer's disease or  
PT cancer -  
XX  
PS Claim 1; Page 31; 318pp; English.  
XX  
CC The present invention relates to new NOX polypeptides. The invention is  
CC useful for creating or preventing a NOX-associated disorder such as  
CC cardiomyopathy or atherosclerosis, where the disorder is related to cell  
CC signal processing and metabolic pathway modulation in a subject,  
CC preferably human. The invention is also useful for treating metabolic  
CC disorders (e.g. obesity), diabetes, infectious disease, neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), immune  
CC disorders, haematopoietic disorders and various cancers. The molecules of  
CC the invention are also useful for treating or preventing cirrhosis,  
CC pancreatitis, learning and memory defects, infertility, congenital heart  
CC defects, acne, hair growth, pigmentation disorders, endocrine disorders,  
CC respiratory disease, gastro-intestinal diseases, reproductive, health,  
CC neurological diseases, bone marrow transplantation, endocrine diseases,  
CC allergy and inflammation, neurological disorders, urinary system  
CC disorders, neuropsychiatric disorders and age-related disorders.  
CC The present amino acid sequence represents a NOX protein of the  
CC invention.  
XX  
SQ Sequence 299 AA;  
XX  
Query Match 89.2%; Score 1540; DB 23; Length 299;  
Best Local Similarity 90.6%; Pred. No. 4e-137;  
Matches 298; Conservative 0; Mismatches 1; Indels 30; Gaps 1;  
QY 1 MLPLLLGLLPACWALGPTPGSSSLRASFSAARTPLGSEMAVTFDKYVNIIGD 60  
DB 1 MLPLLLGLLPACWALGPTPGSSSLRASFSAARTPLGSEMAVTFDKYVNIIGD 60  
QY 61 FDVATGQFRCPVPGAYFFSFTAGKAPHKSLSVMLVRNDEVQALAFDEQRPGARRASQ 120  
DB 61 FDVATGQFRCPVPGAYFFSFTAGKAPHKSLSVMLVRNDEVQALAFDEQRPGARRASQ 120  
QY 121 SAMLOLDYGTWVRLHGAAPHYALGAPGATFSGYLVVADADADAPARPPAPPPRRSAFS 180  
DB 121 SAMLOLDYGTWVRLHGAAPHYALGAPGATFSGYLVVADADADAPARPPAPPPRRSAFS 180  
QY 181 AARTSLVGSAGCPGPRHQPPLAFTEFVNIIGDGFDAAGVRCRLPGAYFFSFTLGLPR 240  
DB 181 AARTSLVGSAGCPGPRHQPPLAFTEFVNIIGDGFDAAGVRCRLPGAYFFSFTLGLPR 240  
QY 241 KTLISVKLMKNRDEVQAMITYDDGASRRRMOSSQSVMLALRRGDVAWLSSHHDGYGAYSNH 300  
DB 241 KTLISVKLMKNRDEVQAMITYDDGASRRRMOSSQSVMLALRRGDVAWLSSHHDGYGAYSNH 300  
QY 271 GKYYTFSGFLVYPDLAPAPPGIGASL 329  
DB 271 GKYYTFSGFLVYPDLAPAPPGIGASL 329  
RESULT 6  
AAB61466  
ID AAB61466 standard; protein; 334 AA.  
XX  
AC AAB61466;  
XX  
DT 04-APR-2001 (first entry)  
XX  
DE Human MANGO 245 mature protein.  
XX  
KW TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
KW autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
KW pancreatic; skeletal; muscle.  
XX  
OS Homo sapiens.  
XX  
PN WO200100672-A1.  
XX  
PD 04-JAN-2001.

XX 29-JUN-2000; 2000WO-US18184.  
XX 29-JUN-1999; 99US-0342687.  
XX (MILL-) MILLENNIUM PHARM INC.  
XX Holtzman DA, Barnes TM, Frazer CC, Sharp JD;  
XX WPI; 2001-050127/06.  
XX  
XX Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
XX TANGO 245 proteins, useful in the treatment of inflammatory diseases  
XX (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
XX disorders (e.g. jaundice) -  
XX  
XX Disclosure; Page 242; 262pp; English.  
XX  
XX The present invention relates to cDNAs encoding TANGO 244,  
XX TANGO 246, TANGO 275, TANGO 300 and TANGO 245 proteins.  
XX The nucleic acids, proteins and protein modulators are useful for  
XX treating colonic disorders, inflammatory diseases, tumors,  
XX renal disorders, liver disorders, lung disorders, autoimmune diseases,  
XX allergic diseases, cardiovascular diseases, brain disorders,  
XX degenerative diseases placental, pancreatic, skeletal and muscle  
XX disorders.  
XX  
XX Sequence 334 AA;  
SQ  
Query Match 86.4%; Score 1491; DB 22; Length 334;  
Best Local Similarity 99.3%; Pred. No. 2e-132;  
Matches 284; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 15 WALGPTPGSGSELRSASFSAARTTPLEGSEMAVTFDKYVYVNIIGDFVATGQRCRVPG 74  
D 1 WALGPTPGSGSELRSASFSAARTTPLEGSEMAVTFDKYVYVNIIGDFVATGQRCRVPG 60  
QY 75 AYFSSFTGKAPKHSLSVMLVRNDEVQALAFDEQRPGARRAASQSMQLDYGDTVWL 134  
D 61 AYFSSFTGKAPKHSLSVMLVRNDEVQALAFDEQRPGARRAASQSMQLDYGDTVWL 120  
QY 135 RLHGAAPHYALGAPATPSGYIVADADADAPARPPAPRPPERSAFSAARTSLVGSNAP 194  
D 121 RLHGAAPHYALGAPATPSGYIVADADADAPARPPAPRPPERSAFSAARTSLVGSNAP 180  
QY 195 GPRHQPPLAFDTEFNIGDFFDAAGVPRCLPGAYFFSFTGKLPKRTLSVKLMKNRDEV 254  
D 181 GPRHQPPLAFDTEFNIGDFFDAAGVPRCLPGAYFFSFTGKLPKRTLSVKLMKNRDEV 240  
QY 255 QAMTYDDGASRRRMOQSVMALRRGDVWLLSHDHGCGAYSNH 300  
D 241 QAMTYDDGASRRRMOQSVMALRRGDVWLLSHDHGCGAYSNH 286  
RESULT 7  
ABG70384 ID ABG70384 standard; Protein; 284 AA.  
XX  
XX ABG70384;  
AC  
XX  
XX 05-NOV-2002 (first entry)  
DE  
XX Adipocyte complement-related C1q Tumour Necrosis Factor-like protein #2.  
XX  
XX Human; NOVX; NOVX-associated disorder; cardiomyopathy; atherosclerosis;  
XX cell signal processing; metabolic pathway modulation; metabolic disorder;  
XX obesity; diabetes; infectious disease; neurodegenerative disorder; acne;  
XX Alzheimer's disease; Parkinson's disease; immune disorder; cancer;  
XX haematopoietic disorder; cirrhosis; pancreatitis; learning defect;  
XX memory defect; infertility; congenital heart defect; hair growth;  
XX implantation disorder; endocrine disorder; respiratory disease; health;  
XX gastro-intestinal disease; reproductive; neurological disease;  
XX bone marrow transplantation; endocrine disease; allergy; inflammation;

XX nephrological disorder; urinary system disorder; age-related disorder;  
XX neuropsychiatric disorder; EGF-related protein; SCUB1; TEN-M4;  
XX adipocyte complement-related C1q tumour necrosis factor; cut at first;  
XX beta adrenergic receptor kinase; EphA6/ehk-2; glucose transporter;  
XX type 1a membrane sushi-containing domain; butyrophillin;  
XX type 1a membrane sushi domain containing.  
XX  
XX Homo sapiens.  
XX  
XX WO200257453-A2.  
XX  
XX 25-JUL-2002.  
XX  
XX 19-DEC-2001; 2001WO-US50331.  
XX  
XX 19-DEC-2000; 2000US-265704P.  
XX 20-DEC-2000; 2000US-257314P.  
XX 02-MAY-2001; 2001US-268153P.  
XX 29-MAY-2001; 2001US-294075P.  
XX 24-JUL-2001; 2001US-307506P.  
XX 10-AUG-2001; 2001US-311590P.  
XX 10-AUG-2001; 2001US-311613P.  
XX 29-AUG-2001; 2001US-315617P.  
XX 14-SEP-2001; 2001US-322358P.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Gangoli EA, Patraja M, Vernet CM, Maljankar UM, Kekuda R;  
XX Stone DJ, Anderson D, Shinkete RA, Burgess CE, Zernhusen BD, Liu X;  
XX Spytek KA, Caeman SJ, Boldog FL, Smithson G, Li L, Ji W;  
XX WPI; 2002-590744/63.  
XX N-PSDB; ABS52096.  
XX  
XX Novel isolated NOVX polypeptide useful for treating cardiomyopathy,  
XX atherosclerosis, metabolic disorders, diabetes, obesity, infectious  
XX disease, anorexia, neurodegenerative disorders, Alzheimer's disease or  
XX cancer -  
XX  
XX Claim 1; Page 30; 318pp; English.  
XX  
XX The present invention relates to new NOVX polypeptides. The invention is  
XX useful for treating or preventing a NOVX-associated disorder such as  
XX cardiomyopathy or atherosclerosis, where the disorder is related to cell  
XX signal processing and metabolic pathway modulation in a subject.  
XX preferably human. The invention is also useful for treating metabolic  
XX disorders (e.g. obesity), diabetes, infectious disease, neurodegenerative  
XX disorders (e.g. Alzheimer's disease, Parkinson's disease), immune  
XX disorders, haematopoietic disorders and various cancers. The molecules of  
XX the invention are also useful for treating or preventing cirrhosis,  
XX pancreatitis, learning and memory defects, infertility, congenital heart  
XX defects, acne, hair growth, pigmentation disorders, endocrine disorders,  
XX respiratory disease, gastro-intestinal diseases, reproductive, health,  
XX neurological diseases, bone marrow transplantation, endocrine diseases,  
XX allergy and inflammation, nephrological disorders, urinary system  
XX disorders, neuropsychiatric disorders and age-related disorders.  
XX The present amino acid sequence represents a NOVX protein of the  
XX invention.  
XX  
XX Sequence 284 AA;  
SQ  
Query Match 83.7%; Score 1445.5; DB 23; Length 284;  
Best Local Similarity 86.0%; Pred. No. 3.2e-128;  
Matches 283; Conservative 0; Mismatches 1; Indels 45; Gaps 1;  
QY 1 MLDPLLLGLLGPAAQWALGPTPGSGSELRSASFSAARTTPLEGSEMAVTFDKYVYVNIIGD 60  
D 1 MLDPLLLGLLGPAAQWALGPTPGSGSELRSASFSAARTTPLEGSEMAVTFDKYVYVNIIGD 60  
QY 61 FVATGQRCRVGAGVAFSFTGKAPKHSLSVMLVRNDEVQALAFDEQRPGARRAASQ 120  
D 61 FVATGQRCRVGAGVAFSFTGKAPKHSLSVMLVRNDEVQALAFDEQRPGARRAASQ 120

QY 121 SAMQLDYGDTVWMLRLHGAAPHYALGAPGATFSGYLVYADADADAPARCPAPPEPRSAFS 180  
DB 121 SAMQLDYGDTVWMLRLHGAAPHYALGAPGATFSGYLVYADADADADAPARCPAPPEPRSAFS 160  
QY 181 AARTSLVSGSDAGPGRHQPPLAFDTEFVNIGSDPFAAGVPRCRLPGAYFFSFTLGKLP 240  
DB 161 -----EFVNIGSDPFAAGVPRCRLPGAYFFSFTLGKLP 195  
QY 241 KTLGVKLMKNRDEVQAMITYDDGASRRRMOGOSVWMLALRRGDVWLLSHDHGYGAYSNH 300  
DB 196 KTLGVKLMKNRDEVQAMITYDDGASRRRMOGOSVWMLALRRGDVWLLSHDHGYGAYSNH 255  
QY 301 GKYTFSGFLVYPDLAPAPPGGLGASELL 329  
DB 256 GKYTFSGFLVYPDLAPAPPGGLGASELL 284

## RESULT 8

AAB61488 ID AAB61488 standard; protein; 199 AA.

AC AAB61488;  
XX  
DT 04-APR-2001 (first entry)  
XX  
DE Murine MANGO 245 protein.  
XX  
KW TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
KM autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
KN pancreatic; skeletal; muscle.  
XX  
OS Mus musculus.  
XX  
PN WO200100672-A1.  
XX  
PD 04-JAN-2001.  
XX  
PE 29-JUN-2000; 2000WO-US18184.  
XX  
PF 29-JUN-1999; 99US-0342687.  
XX  
PR 29-JUN-1999; 99US-0342687.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;  
DR WPI; 2001-050127/06.  
XX  
PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
XX MANGO 245 proteins, useful in the treatment of inflammatory diseases  
XX (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
XX disorders (e.g. jaundice) -  
XX  
PS Claim 1; Fig 29; 262pp; English.  
XX  
CC The present invention relates to cDNAs encoding TANGO 244,  
XX TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
XX The nucleic acids, proteins and protein modulators are useful for  
XX treating colonic disorders, inflammatory diseases, tumors,  
XX renal disorders, liver disorders, lung disorders, autoimmune diseases,  
XX allergic diseases, cardiovascular diseases, brain disorders,  
XX degenerative diseases placental, pancreatic, skeletal and muscle  
XX disorders.  
SQ Sequence 199 AA;

Query Match 51.1%; Score 882; DB 22; Length 199;  
Best Local Similarity 87.2%; Pred. No. 3.5e-75;  
Matches 177; Conservative 4; Mismatches 18; Indels 4; Gaps 3;  
DB 1 MLLPLGLGPPACALGPPA-GPGSSELRSARSAARTTPLEGTSMAVTFDKYVNIIGSD 59

QY 61 FDVATGCRRCRVPAAYFFSFTAGKAPKKSLSVWLVRNREVOALAFDQRRPGRRAASQ 120  
DB 60 FDAATGFRRCRVPAAYFFSFTAGKAPKKSLSVWLVRNREVOALAFDQRRPGRRAASQ 119  
QY 121 SAMQLDYGDTVWMLRLHGAAPHYALGAPGATFSGYLVYADADADAPARCPAPPEPRSAFS 180  
DB 120 SAMQLDYGDTVWMLRLHGAAPHYALGAPGATFSGYLVYADADADAPARCPAPPEPRSAFS 178  
QY 181 AARTSLVSGSDAGPGRHQPPLAF 203  
DB 179 ARHATWAPNPPRPPRR--LAF 199

## RESULT 9

AAB61479 ID AAB61479 standard; protein; 192 AA.

AC AAB61479;  
XX  
DT 04-APR-2001 (first entry)  
XX  
DE Human MANGO 245 cytoplasmic domain.  
XX  
KW TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
KM autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
KN pancreatic; skeletal; muscle.  
XX  
OS Homo sapiens.  
XX  
PN WO200100672-A1.  
XX  
PD 04-JAN-2001.  
XX  
PE 29-JUN-2000; 2000WO-US18184.  
XX  
PF 29-JUN-1999; 99US-0342687.  
XX  
PR 29-JUN-1999; 99US-0342687.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;  
DR WPI; 2001-050127/06.  
XX  
PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
XX MANGO 245 proteins, useful in the treatment of inflammatory diseases  
XX (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
XX disorders (e.g. jaundice) -  
XX  
PS Disclosure; Page 247; 262pp; English.  
XX  
CC The present invention relates to cDNAs encoding TANGO 244,  
XX TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
XX The nucleic acids, proteins and protein modulators are useful for  
XX treating colonic disorders, inflammatory diseases, tumors,  
XX renal disorders, liver disorders, lung disorders, autoimmune diseases,  
XX allergic diseases, cardiovascular diseases, brain disorders,  
XX degenerative diseases placental, pancreatic, skeletal and muscle  
XX disorders.  
SQ Sequence 192 AA;

Query Match 44.0%; Score 760; DB 22; Length 192;  
Best Local Similarity 100.0%; Pred. No. 1.1e-63;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 157 YADADADAPARCPAPPEPRSAARTSLVSGSDAGPGRHQPPLAFDTEFVNIGSDPDA 216  
DB 1 YADADADAPARCPAPPEPRSAARTSLVSGSDAGPGRHQPPLAFDTEFVNIGSDPDA 60  
QY 217 AAGVPRCRLPAAYFFSFTLGKLPKTLGVKLMKNRDEVQAMITYDDGASRRRMOGOSVWL 276  
DB 61 AAGVPRCRLPAAYFFSFTLGKLPKTLGVKLMKNRDEVQAMITYDDGASRRRMOGOSVWL 120

QY 277 ALRRGDAVWLISHDHYGAYSNH 300  
Db 121 ALRRGDAVWLISHDHYGAYSNH 144

RESULT 10  
ABG70383  
ID ABG70383 standard; Protein, 221 AA.  
XX  
AC ABG70383;  
XX  
DT 05-NOV-2002 (first entry)  
XX  
DE Adipocyte complement-related C1q Tumour Necrosis Factor-like protein #1.  
XX  
KW Human; NOX; NOX-associated disorder; cardiomyopathy; atherosclerosis;  
KW cell signal processing; metabolic pathway modulation; metabolic disorder;  
KW obesity; diabetes; infectious disease; neurodegenerative disorder; acne;  
KW Alzheimer's disease; Parkinson's disease; immune disorder; cancer;  
KW haematopoietic disorder; cirrhosis; pancreatitis; learning defect;  
KW memory defect; infertility; congenital heart defect; hair growth;  
KW pigmentary disorder; endocrine disorder; respiratory disease; health;  
KW gastro-intestinal disease; reproductive; neurological disease;  
KW bone marrow transplantation; endocrine disease; allergy; inflammation;  
KW nephrological disorder; urinary system disorder; age-related disorder;  
KW neuropsychiatric disorder; EGF-related protein; SCUB1; TEN-M4;  
KW adipocyte complement-related C1q tumour necrosis factor; cut at first;  
KW beta adrenergic receptor kinase; Epna6/bk-2; glucose transporter;  
KW type Ia membrane sushi-containing domain; butyrophilin;  
XX type Ia membrane sushi domain containing.  
XX  
OS Homo sapiens.  
XX  
PN W0200257453-A2.  
XX  
PD 25-JUL-2002.  
XX  
PF 19-DEC-2001; 2001WO-US50331.  
XX  
PR 19-DEC-2000; 2000US-265704P.  
XX 20-DEC-2000; 2000US-257314P.  
PR 02-MAY-2001; 2001US-288153P.  
PR 29-MAY-2001; 2001US-294075P.  
PR 24-JUL-2001; 2001US-307506P.  
PR 10-AUG-2001; 2001US-311590P.  
PR 10-AUG-2001; 2001US-311613P.  
PR 29-AUG-2001; 2001US-315617P.  
PR 14-SEP-2001; 2001US-322358P.  
XX  
XX (CURAGEN CORP.  
XX  
PI Gangoli EA, Patturajan M, Vernet CAM, Malysankar UM, Kekuda R;  
PI Stone DJ, Anderson D, Shimkets RA, Burgess CE, Zernhusen BD, Liu X;  
PI Szytek KA, Casman SJ, Boldog FL, Smithson G, Li U, Ji W;  
XX WPI; 2002-590744/63.  
DR N-PSDB; ABS52095.  
XX  
XX Novel isolated NOX polypeptide useful for treating cardiomyopathy,  
PT atherosclerosis, metabolic disorders, diabetes, obesity, infectious  
PT disease, anorexia, neurodegenerative disorders, Alzheimer's disease or  
PT cancer.  
XX  
PS Claim 1; Page 28; 318pp; English.  
XX

CC the invention are also useful for treating or preventing cirrhosis,  
CC pancreatitis, learning and memory defects, infertility, congenital heart  
CC defects, acne, hair growth, pigmentation disorders, endocrine disorders,  
CC respiratory disease, gastro-intestinal diseases, reproductive, health,  
CC neurological diseases, bone marrow transplantation, endocrine diseases,  
CC allergy and inflammation, nephrological disorders, urinary system  
CC disorders, neuropsychiatric disorders and age-related disorders.  
CC The present amino acid sequence represents a NOX protein of the  
CC invention.  
XX  
SQ Sequence 221 AA;  
XX  
Query Match 41.0%; Score 707.5; DB 23; Length 221;  
Best Local Similarity 49.5%; Pred. No. 1,3e-58;  
Matches 153; Conservative 1; Mismatches 2; Indels 153; Gaps 1;

QY 21 PGQSSSLRFAFSARTTPEGTSEMAVTFDKYVYVIGDFVATGFRCPGAYFFSF 80  
Db 66 PGQSSSLRFAFSARTTPEGTSEMAVTFDKYVYVIGDFVATGFRCPGAYFFSF 99  
QY 81 TAGKAPKSLVNLVNRDEVQALAFDEORRPGARRAASQAMLQLDYDPTWLRHGAP 140  
Db 100 ----- 99  
QY 141 HYALGAGATFSGYLVYADADADAPARGPAPPEPPSASFAARTSLVGS DAGCPRHQP 200  
Db 100 ----- 99  
QY 201 LAFDTFVYVIGDFPDAAGVRCRLPGAYFFSFTLGLPKRTLSVKLMKRDEVQAMTYD 260  
Db 100 -----VNIGDFPDAAGVRCRLPGAYFFSFTLGLPKRTLSVKLMKRDEVQAMTYD 152  
QY 261 DGASRRRMOQSQSVMLALRRGDAVWLISHDHYGAYSNHGXYITTSGLFVYDLPAPAP 320  
Db 153 DGASRRRMOQSQSVMLALRRGDAVWLISHDHYGAYSNHGXYITTSGLFVYDLPAPAP 212  
QY 321 PGIGASELL 329  
Db 213 PGIGASELL 221

RESULT 11  
AAB61473  
ID AAB61473 standard; Protein, 134 AA.  
XX  
AC AAB61473;  
XX  
DT 04-APR-2001 (first entry)  
XX  
DE Monkey MANGO 245 C1q domain #2.  
XX  
KW TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
KW autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
KW pancreatic; skeletal; muscle.  
XX  
OS Catarrhini sp.  
XX  
PN W0200100672-A1.  
XX  
PD 04-JAN-2001.  
XX  
PF 29-JUN-2000; 2000WO-US18184.  
XX  
PR 29-JUN-1999; 99US-0342687.  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;  
XX WPI; 2001-050127/06.  
DR Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases

FT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
 PT disorders (e.g. jaundice) -  
 PS Disclousure; Fig 28; 262pp; English.  
 XX  
 CC The present invention relates to cDNAs encoding TANGO 244,  
 CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
 CC The nucleic acids, proteins and protein modulators are useful for  
 CC treating colonic disorders, inflammatory diseases, tumors,  
 CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
 CC allergic diseases, cardiovascular diseases, brain disorders,  
 CC degenerative diseases placental, pancreatic, skeletal and muscle  
 CC disorders.  
 CC  
 XX  
 SQ Sequence 134 AA;  
 Query Match 39.6%; Score 683; DB 22; Length 134;  
 Best Local Similarity 97.8%; Pred. No. 1.3e-56;  
 Matches 131; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 178 AFSARTSLVGSDDAGPGRHQPLAFDTEFVNIIGDFDPAAGVFCRPLPGAVFFSFTLGK 237  
 DB 1 AFSARTSLVGSDDAGPGRHQPLAFDTEFVNIIGDFDPAAGVFCRPLPGAVFFSFTLGK 60  
 QY 238 LPRKTLVYLMKNRDEVOAMTYDDGASRRREMOSQVWLALRRGDVWLSSHDDGYGAY 297  
 DB 61 LPRKTLVYLMKNRDEVOAMTYDDGASRRREMOSQVWLALRRGDVWLSSHDDGYGAY 120  
 QY 298 SNHGKXITFSGFLV 311  
 DB 121 SNHGKXITFSGFLV 134  
 RESULT 12  
 ABB11578  
 ID ABB11578 standard; peptide; 127 AA.  
 XX  
 AC ABB11578;  
 XX  
 DT 11-JAN-2002 (first entry)  
 XX  
 DE Human secreted protein homologue, SEQ ID NO:1948.  
 XX  
 KW Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 KW haematopoiesis regulation; tissue growth; immunomodulator; actin;  
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
 KW chronic inflammatory condition; proliferative retinopathy;  
 KW atherosclerosis; coronary heart disease; arterial ischaemia;  
 KW bone disorder; osteoporosis; vascular growth disorder;  
 KW tissue regeneration; wound healing; infection; immune disorder;  
 KW cell culture; drug screening; gene therapy; antiinflammatory;  
 KW antiaesthetic; antiarthritic; haemostatic; antiarteriosclerotic;  
 KW cytotetic; osteopathic; vasotropic; cardiant; vitrucide; antibacterial;  
 KW antifungal; vulnerary; antiulcer.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200157188-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PE 05-FEB-2001; 2001WO-US03800.  
 XX  
 PR 03-FEB-2000; 2000US-0496914.  
 XX 27-APR-2000; 2000US-0560875.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX  
 DR WPI; 2001-457740/49.

DR N-PSDB; ABA08822.  
 XX  
 PT Human proteins and DNA encoding sequences useful for preventing,  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer -  
 XX  
 XX Claim 20; Page 214-215; 1963pp; English.  
 XX  
 SS Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 CC invention also relates to vectors and recombinant host cells comprising a  
 CC nucleotide of the invention, methods of producing the novel polypeptides,  
 CC antibodies against the polypeptides, methods of detecting the nucleotides,  
 CC or polypeptides in a sample, and methods of identifying compounds which  
 CC bind to polypeptides of the invention. Although novel, many of the  
 CC polypeptides of the invention have homology to known proteins, thereby  
 CC giving an insight into their probable biological activities, and hence  
 CC potential therapeutic applications. The polypeptides of the invention may  
 CC have various activities, including cytokine, cell proliferation or cell  
 CC differentiation activities; stem cell growth factor activity;  
 CC haematopoiesis regulatory activity; tissue growth activity;  
 CC immunomodulatory activity; activin- or inhibin-related activities;  
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 CC thrombolytic activities; receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC Depending on their biological activities, polypeptides and nucleotides of  
 CC the invention are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC vascular growth. Polypeptides involved with tissue regeneration and  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders.  
 CC Polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells  
 CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a novel human  
 CC polypeptide of the invention.  
 CC  
 SQ Sequence 127 AA;  
 Query Match 38.5%; Score 664; DB 22; Length 127;  
 Best Local Similarity 99.2%; Pred. No. 7.8e-55;  
 Matches 126; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 203 FDFEFVNIIGDFDPAAGVFCRPLPGAVFFSFTLGKLPKTLVYLMKNRDEVOAMTYDDG 262  
 DB 1 FDFEFVNIIGDFDPAAGVFCRPLPGAVFFSFTLGKLPKTLVYLMKNRDEVOAMTYDDG 60  
 QY 263 ASRRREMOSQVWLALRRGDVWLSSHDDGYGAYSNHGKXITFSGFLVYPDLAAPAPG 322  
 DB 61 SSRRREMOSQVWLALRRGDVWLSSHDDGYGAYSNHGKXITFSGFLVYPDLAAPAPG 120  
 QY 323 LGASELL 329  
 DB 121 LGASELL 127  
 RESULT 13  
 AAB61477  
 ID AAB61477 standard; protein; 125 AA.  
 XX  
 AC AAB61477;  
 XX  
 DT 04-APR-2001 (first entry)  
 XX



DE Human MANGO 245 extracellular domain.  
XX  
KW TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
KW autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
KW pancreatic; skeletal; muscle.  
XX  
OS Homo sapiens.  
XX  
PN WO200100672-A1.  
XX  
PD 04-JAN-2001.  
XX  
PF 29-JUN-2000; 2000WO-US18184.  
XX  
PR 29-JUN-1999; 99US-0342687.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;  
XX  
DR WPI; 2001-050127/06.  
XX  
PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
PT disorders (e.g. jaundice) -  
XX  
PS Disclosure; Page 246; 262pp; English.  
XX  
CC The present invention relates to cDNAs encoding TANGO 244,  
CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
CC The nucleic acids, proteins and protein modulators are useful for  
CC treating colonic disorders, inflammatory diseases, tumors,  
CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
CC allergic diseases, cardiovascular diseases, brain disorders,  
CC degenerative diseases placental, pancreatic, skeletal and muscle  
CC disorders.  
XX  
SQ Sequence 125 AA;  
XX  
Query Match 37.7%; Score 650; DB 22; Length 125;  
Best Local Similarity 100.0%; Pred. NO. 1.6e-53;  
Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 15 WALGPTGPGSSSELRSASFSAARTTPLEGTSMAVTFDKYYVNIIGDPDVATGQRCRVPG 74  
DB 1 WALGPTGPGSSSELRSASFSAARTTPLEGTSMAVTFDKYYVNIIGDPDVATGQRCRVPG 60  
QY 75 AYFSPFTAGKAPHKSLSYMLVNRNDEVQALAFDEQRPGARRAASOSAMLQLDYGTIVWL 134  
DB 61 AYFSPFTAGKAPHKSLSYMLVNRNDEVQALAFDEQRPGARRAASOSAMLQLDYGTIVWL 120  
QY 135 RLHGA 139  
DB 121 RLHGA 125  
XX  
RESULT 14  
ID AAB61469  
XX AAB61469 standard; protein; 126 AA.  
XX  
AC AAB61469;  
XX  
DT 04-APR-2001 (first entry)  
XX  
DE Human MANGO 245 C1q domain #1.  
XX  
KW TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
KW autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
KW pancreatic; skeletal; muscle.  
XX  
OS Homo sapiens.  
XX

PN WO200100672-A1.  
XX  
PD 04-JAN-2001.  
XX  
PF 29-JUN-2000; 2000WO-US18184.  
XX  
PR 29-JUN-1999; 99US-0342687.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;  
XX  
DR WPI; 2001-050127/06.  
XX  
PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
PT disorders (e.g. jaundice) -  
XX  
PS Disclosure; Fig 26; 262pp; English.  
XX  
CC The present invention relates to cDNAs encoding TANGO 244,  
CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
CC The nucleic acids, proteins and protein modulators are useful for  
CC treating colonic disorders, inflammatory diseases, tumors,  
CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
CC allergic diseases, cardiovascular diseases, brain disorders,  
CC degenerative diseases placental, pancreatic, skeletal and muscle  
CC disorders.  
XX  
SQ Sequence 126 AA;  
XX  
Query Match 37.4%; Score 646; DB 22; Length 126;  
Best Local Similarity 99.2%; Pred. NO. 3.9e-53;  
Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 31 AFSARTTPLEGTSMAVTFDKYYVNIIGDPDVATGQRCRVGAYFFSTAGKAPHKSL 90  
DB 1 AFSARTTPLEGTSMAVTFDKYYVNIIGDPDVATGQRCRVGAYFFSTAGKAPHKSL 60  
QY 91 SVMLVNRNDEVQALAFDEQRPGARRAASOSAMLQLDYGTIVWLRLHGAHYALGARGAT 150  
DB 61 SVMLVNRNDEVQALAFDEQRPGARRAASOSAMLQLDYGTIVWLRLHGAHYALGARGAT 120  
QY 151 FSGYLV 156  
DB 121 FSGYLV 126  
XX  
RESULT 15  
ID AAB61489  
XX AAB61489 standard; protein; 126 AA.  
XX  
AC AAB61489;  
XX  
DT 04-APR-2001 (first entry)  
XX  
DE Murine MANGO 245 C1q domain.  
XX  
KW TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
KW autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
KW pancreatic; skeletal; muscle.  
XX  
OS Mus musculus.  
XX  
PN WO200100672-A1.  
XX  
PD 04-JAN-2001.  
XX  
PF 29-JUN-2000; 2000WO-US18184.  
XX  
PR 29-JUN-1999; 99US-0342687.  
XX

PA (MILL-) MILLENNIUM PHARM INC.

XX Holtzman DA, Barnes TM, Fraser CC, Sharp JD;

XX WPI; 2001-050127/06.

XX  
PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
PT disorders (e.g. jaundice) -

XX  
PS Disclosure: Page 260-261; 262pp; English.

XX  
CC The present invention relates to cDNAs encoding TANGO 244,  
CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
CC The nucleic acids, proteins and protein modulators are useful for  
CC treating colonic disorders, inflammatory diseases, tumors,  
CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
CC allergic diseases, cardiovascular diseases, brain disorders,  
CC degenerative diseases placental, pancreatic, skeletal and muscle  
CC disorders.

XX  
SQ Sequence 126 AA:

Query Match 37.0%; Score 639; DB 22; Length 126;  
Best Local Similarity 96.8%; Pred. No. 1.8e-52;  
Matches 122; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 31 AFSARTTPLECTSEMAVTGDKVYVNIIGDPPVATGOFRCRVPGAYFFSFTAGKAPHKSL 90  
DB 1 AFSARTTPLECTSEMAVTGDKVYVNIIGDPPVATGOFRCRVPGAYFFSFTAGKAPHKSL 60  
OY 91 SVMIVRNREDEVQALAFDEQRRPGARRAASQSAMQLDYGDTVWRLHGA PHYALGAPGAT 150  
DB 61 SVMIVRNREDEVQALAFDEQRRPGARRAASQSAMQLDYGDTVWRLHGA PHYALGAPGAT 120  
OY 151 FSGYLV 156  
DB 121 FSGYLV 126

Search completed: January 12, 2004, 08:15:34  
Job time : 51.5798 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: January 12, 2004, 08:15:44 ; Search time 40.0579 Seconds

(without alignments)  
1655.023 Million cell updates/sec

Title: US-10-085-167-2

Perfect score: 1726  
Sequence: 1 MRLPLGLGPAACWALGPT.....LYPDLAPAPPGUGSELL 329

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747907 seqs, 201509753 residues

Total number of hits satisfying chosen parameters: 747907

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*

2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*

3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*

4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*

5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*

6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*

7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*

8: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*

9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*

10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*

11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*

12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*

13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*

14: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*

15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*

16: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*

17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*

18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1726	100.0	329	US-10-085-167-2	Sequence 2, App1
2	1720	99.7	329	US-10-236-055A-14	Sequence 14, App1
3	1587	91.9	326	US-10-236-055A-16	Sequence 16, App1
4	237.5	13.8	252	US-09-993-737-52	Sequence 52, App1
5	230	13.3	202	US-10-091-458-43	Sequence 43, App1
6	230	13.3	420	US-10-236-055A-2	Sequence 2, App1
7	230	13.3	746	US-09-738-973-185	Sequence 185, App
8	230	13.3	746	US-09-854-133-185	Sequence 185, App
9	230	13.3	746	US-10-144-649A-185	Sequence 185, App
10	230	13.3	908	US-10-309-422-10	Sequence 10, App1
11	230	13.3	909	US-10-309-422-22	Sequence 22, App1
12	230	13.3	957	US-10-309-422-14	Sequence 14, App1
13	230	13.3	958	US-10-309-422-26	Sequence 26, App1
14	230	13.3	992	US-10-309-422-18	Sequence 18, App1
15	230	13.3	993	US-10-309-422-30	Sequence 30, App1

16	230	13.3	1042	12	US-10-309-422-8	Sequence 8, App1
17	230	13.3	1043	12	US-10-309-422-20	Sequence 20, App1
18	230	13.3	1091	12	US-10-309-422-12	Sequence 12, App1
19	230	13.3	1092	12	US-10-309-422-24	Sequence 24, App1
20	230	13.3	1126	12	US-10-309-422-16	Sequence 16, App1
21	230	13.3	1127	12	US-10-309-422-28	Sequence 28, App1
22	226	13.1	409	12	US-10-236-055A-4	Sequence 4, App1
23	220.5	12.8	281	10	US-09-911-176B-44	Sequence 44, App1
24	220.5	12.8	281	12	US-10-360-186-44	Sequence 44, App1
25	220.5	12.8	281	12	US-10-392-531-44	Sequence 44, App1
26	220.5	12.8	281	12	US-10-392-706-44	Sequence 44, App1
27	220.5	12.8	281	12	US-10-385-015-5	Sequence 5, App1
28	220.5	12.8	281	15	US-10-180-762-44	Sequence 44, App1
29	220.5	12.8	281	15	US-10-241-258-44	Sequence 44, App1
30	216	12.5	281	9	US-09-768-926-46	Sequence 46, App1
31	215	12.5	281	9	US-09-765-205-24	Sequence 24, App1
32	215	12.5	281	9	US-09-866-028-78	Sequence 78, App1
33	215	12.5	281	10	US-09-944-449-78	Sequence 78, App1
34	215	12.5	281	10	US-09-944-457-78	Sequence 78, App1
35	215	12.5	281	10	US-09-944-862-78	Sequence 78, App1
36	215	12.5	281	10	US-09-945-587-78	Sequence 78, App1
37	215	12.5	281	10	US-09-945-015-78	Sequence 78, App1
38	215	12.5	281	10	US-09-944-396-78	Sequence 78, App1
39	215	12.5	281	10	US-09-944-097-78	Sequence 78, App1
40	215	12.5	281	10	US-09-944-432-78	Sequence 78, App1
41	215	12.5	281	10	US-09-943-762-78	Sequence 78, App1
42	215	12.5	281	10	US-09-944-654-78	Sequence 78, App1
43	215	12.5	281	10	US-09-943-851A-78	Sequence 78, App1
44	215	12.5	281	10	US-09-944-413-78	Sequence 78, App1
45	215	12.5	281	10	US-09-911-176B-2	Sequence 2, App1

## ALIGNMENTS

RESULT 1  
US-10-085-167-2  
; Sequence 2, Application US/10085167  
; Publication No. US20030170781A1  
; GENERAL INFORMATION:  
; APPLICANT: Hollway, James L.  
; TITLE OF INVENTION: SECRETED PROTEIN ZACRP4  
; FILE REFERENCE: 99-29  
; CURRENT APPLICATION NUMBER: US/10/085,167  
; CURRENT FILING DATE: 2002-02-27  
; PRIOR APPLICATION NUMBER: 60/141,928  
; PRIOR FILING DATE: 1999-07-01  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 329  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-085-167-2

Query Match 100.0%; Score 1726; DB 12; Length 329;

Best Local Similarity 100.0%; Pred. No. 3 4e-154;

Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MRLPLGLGPAACWALGPTPGSSSELSRFAAARTTPLEGSEMAVTFDKYVNI	60
DB	1	MRLPLGLGPAACWALGPTPGSSSELSRFAAARTTPLEGSEMAVTFDKYVNI	60
QY	61	FDVATQFRVRVGAFFSTTACKAPKKSIVLVNRRDEVQALAFDEQRRPARASQ	120
DB	61	FDVATQFRVRVGAFFSTTACKAPKKSIVLVNRRDEVQALAFDEQRRPARASQ	120
QY	121	SAMQLDYGDTWRLRIGAPHYVALGAPATFSYLVYADADADAPARGPAPPEPSAFS	180
DB	121	SAMQLDYGDTWRLRIGAPHYVALGAPATFSYLVYADADADAPARGPAPPEPSAFS	180
QY	181	AARTSLVGSADGPGRRHOPAFDTFVNIGDFDAAGVFRCLPGAYFFSFTLGLKLR	240

Db 181 AARTSLVGSADGPGPRHQPPLAFDTEFVNIGDGFPAAGVFCRLPGAYFFSFTLGKLP 240  
QY 241 KTLSTVKLMKNRDEVQAMTYDDGASRRREMOSQSVMLALRRGDVWLLSHDHGCGAYSNH 300  
Db 241 KTLSTVKLMKNRDEVQAMTYDDGASRRREMOSQSVMLALRRGDVWLLSHDHGCGAYSNH 300  
QY 301 GKTYTFSGFLVYPDLAPAPPGIGASELL 329  
Db 301 GKTYTFSGFLVYPDLAPAPPGIGASELL 329

RESULT 2  
US-10-236-055A-14  
; Sequence 14, Application US/10236055A  
; Publication No. US20030134328A1  
; GENERAL INFORMATION:  
; APPLICANT: Basham, Beth E.  
; APPLICANT: Forsythe, Ian  
; APPLICANT: Gorman, Daniel M.  
; APPLICANT: Mattson, Jeanine  
; APPLICANT: Moshrefi, Mehrdad  
; APPLICANT: Parham, Christi  
; TITLE OF INVENTION: MAMMALIAN GENES, RELATED REAGENTS  
; FILE REFERENCE: DX01343K  
; CURRENT APPLICATION NUMBER: US/10/236,055A  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/317,988  
; PRIOR FILING DATE: 2001-09-06  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 14  
; LENGTH: 329  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-236-055A-14

Query Match 99.7%; Score 1720; DB 12; Length 329;  
Best Local Similarity 99.7%; Pred. No. 1.3e-153;  
Matches 328; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLLPLILGLGPAACWALGPTPGSSSELSAFSAARTTPLEGTSMAVTPDKYVNIIGD 60  
Db 1 MLLPLILGLGPAACWALGPTPGSSSELSAFSAARTTPLEGTSMAVTPDKYVNIIGD 60  
QY 61 FDVATGQFRCRVPGAYFFSFTAGKAPKHSLSVMLVNRNDEVQALAFDEQRPGARRAASQ 120  
Db 61 FDVATGQFRCRVPGAYFFSFTAGKAPKHSLSVMLVNRNDEVQALAFDEQRPGARRAASQ 120  
QY 121 SAMLQLDYGDVTWMLRHGAHYALGAPATFSGYLVYADADADAPARGPAPPEPRSAFS 180  
Db 121 SAMLQLDYGDVTWMLRHGAHYALGAPATFSGYLVYADADADAPARGPAPPEPRSAFS 180  
QY 181 AARTSLVGSADGPGPRHQPPLAFDTEFVNIGDGFPAAGVFCRLPGAYFFSFTLGKLP 240  
Db 181 AARTSLVGSADGPGPRHQPPLAFDTEFVNIGDGFPAAGVFCRLPGAYFFSFTLGKLP 240  
QY 241 KTLSTVKLMKNRDEVQAMTYDDGASRRREMOSQSVMLALRRGDVWLLSHDHGCGAYSNH 300  
Db 241 KTLSTVKLMKNRDEVQAMTYDDGASRRREMOSQSVMLALRRGDVWLLSHDHGCGAYSNH 300  
QY 301 GKTYTFSGFLVYPDLAPAPPGIGASELL 329  
Db 301 GKTYTFSGFLVYPDLAPAPPGIGASELL 329

RESULT 3  
US-10-236-055A-16  
; Sequence 16, Application US/10236055A  
; Publication No. US20030134328A1  
; GENERAL INFORMATION:  
; APPLICANT: Basham, Beth E.  
; APPLICANT: Forsythe, Ian

APPLICANT: Gorman, Daniel M.  
APPLICANT: Mattson, Jeanine  
APPLICANT: Moshrefi, Mehrdad  
APPLICANT: Parham, Christi  
TITLE OF INVENTION: MAMMALIAN GENES, RELATED REAGENTS  
FILE REFERENCE: DX01343K  
CURRENT APPLICATION NUMBER: US/10/236,055A  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/317,988  
PRIOR FILING DATE: 2001-09-06  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 16  
LENGTH: 326  
TYPE: PRT  
ORGANISM: Mus musculus  
US-10-236-055A-16

Query Match 91.9%; Score 1587; DB 12; Length 326;  
Best Local Similarity 94.2%; Pred. No. 4.2e-141;  
Matches 309; Conservative 2; Mismatches 15; Indels 2; Gaps 2;

QY 1 MLLPLILGLGPAACWALGPTPGSSSELSAFSAARTTPLEGTSMAVTPDKYVNIIGD 60  
Db 1 MLLPLILGLGPAACWALGPA-GPSSSELSAFSAARTTPLEGTSMAVTPDKYVNIIGD 59  
QY 61 FDVATGQFRCRVPGAYFFSFTAGKAPKHSLSVMLVNRNDEVQALAFDEQRPGARRAASQ 120  
Db 61 FDVATGQFRCRVPGAYFFSFTAGKAPKHSLSVMLVNRNDEVQALAFDEQRPGARRAASQ 119  
QY 121 SAMLQLDYGDVTWMLRHGAHYALGAPATFSGYLVYADADADAPARGPAPPEPRSAFS 180  
Db 120 SAMLQLDYGDVTWMLRHGAHYALGAPATFSGYLVYADADADAPARGPAPPEPRSAFS 178  
QY 181 AARTSLVGSADGPGPRHQPPLAFDTEFVNIGDGFPAAGVFCRLPGAYFFSFTLGKLP 240  
Db 179 AARTSLVGSADGPGPRHQPPLAFDTEFVNIGDGFPAAGVFCRLPGAYFFSFTLGKLP 238  
QY 241 KTLSTVKLMKNRDEVQAMTYDDGASRRREMOSQSVMLALRRGDVWLLSHDHGCGAYSNH 300  
Db 239 KTLSTVKLMKNRDEVQAMTYDDGASRRREMOSQSVMLALRRGDVWLLSHDHGCGAYSNH 298  
QY 301 GKTYTFSGFLVYPDLAPAPPGIGASELL 328  
Db 299 GKTYTFSGFLVYPDLAPAPPGIGASELL 326

RESULT 4  
US-09-893-737-52  
; Sequence 52, Application US/09893737  
; Patent No. US20020110855A1  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; APPLICANT: Preenell, Scott R.  
; TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS  
; FILE REFERENCE: 00-41  
; CURRENT APPLICATION NUMBER: US/09/893,737  
; CURRENT FILING DATE: 2001-06-28  
; PRIOR APPLICATION NUMBER: US 60/215,446  
; PRIOR FILING DATE: 2000-06-30  
; NUMBER OF SEQ ID NOS: 329  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 52  
; LENGTH: 252  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-893-737-52

Query Match 13.8%; Score 237.5; DB 10; Length 252;  
Best Local Similarity 43.0%; Pred. No. 3.7e-14;  
Matches 65; Conservative 12; Mismatches 61; Indels 13; Gaps 7;

QY 168 GPPAPPEPR--SAFSAARTSLVGSADGPGPRHQPPLAFDTEFVNIGDGFPAAGVFCRL 225

Db 106 GPFQACRAAFAATSVGREGHSSD-----HFOAVPFDTELVMIDGAFDLAAGHFLCTV 160  
Qy 226 FGAVFSTTLKLPKTLISVLMKNNRDEVQAMTYDDGASRRRQSQSVMALRRGDAVW 285  
Db 161 FGAVFSLINHTWMTWKETYHLHIMNR--RPAVLVAQ--PSEKSVNQAGSLMLLLAAGDAVW 218  
Qy 286 --LLSHHDGAGAYSNNCK-YITFSGFLVYP 313  
Db 219 VRMFQDRDN-AIYGEHGLDYITFSGHLVKP 248

RESULT 5  
US-10-091-458-43  
Sequence 43, Application US/10091458  
Publication No. US20030068627A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PTZ06C1  
CURRENT APPLICATION NUMBER: US/10/091,458  
PRIOR FILING DATE: 2002-03-07  
PRIOR APPLICATION NUMBER: 09/764,900  
PRIOR FILING DATE: 2001-01-17  
PRIOR APPLICATION NUMBER: 60/179,065  
PRIOR FILING DATE: 2000-01-31  
PRIOR APPLICATION NUMBER: 60/180,628  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: 60/214,886  
PRIOR FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 60/217,487  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/225,758  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/220,963  
PRIOR FILING DATE: 2000-07-26  
PRIOR APPLICATION NUMBER: 60/217,496  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/225,447  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/218,290  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: 60/225,757  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/226,868  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/216,647  
PRIOR FILING DATE: 2000-07-07  
PRIOR APPLICATION NUMBER: 60/225,267  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/216,880  
PRIOR FILING DATE: 2000-07-07  
PRIOR APPLICATION NUMBER: 60/225,270  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/251,869  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/235,834  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: 60/234,274  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: 60/234,223  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: 60/228,924  
PRIOR FILING DATE: 2000-08-30  
PRIOR APPLICATION NUMBER: 60/224,518  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/236,369  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/224,519  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/220,964  
PRIOR FILING DATE: 2000-07-26  
PRIOR APPLICATION NUMBER: 60/241,809

PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/249,299  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/236,327  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/241,785  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/244,617  
PRIOR FILING DATE: 2000-11-01  
PRIOR APPLICATION NUMBER: 60/225,268  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/236,368  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/251,856  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/251,868  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/229,344  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/234,997  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: 60/229,343  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,345  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,287  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,513  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: 60/231,413  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/229,509  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: 60/236,367  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/237,039  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,038  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/236,370  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/236,802  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,037  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,040  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/240,960  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/239,935  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 60/239,937  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 60/241,787  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/246,474  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/246,532  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/249,216  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,210  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/226,681  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/225,759  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/225,213  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/227,182  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/225,214  
PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/235,836  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: 60/230,438  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/215,135  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: 60/225,266  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/249,218  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,208  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,213  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,212  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,207  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,245  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,244  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,217  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,211  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,215  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,264  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,214  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,297  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/231,242  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/232,081  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/232,080  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/231,414  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/231,244  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/233,064  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/233,063  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232,397  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232,399  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232,401  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/241,808  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241,826  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241,786  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241,221  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/246,475  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/231,243  
PRIOR FILING DATE: 2000-09-08

Query Match 13.3%; Score 230; DB 15; Length 202;  
Best Local Similarity 38.4%; Pred. No. 1,4e-13;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

Qy 163 DAPARGP-----PAPPEPSAFSAARTSLVSGDAGPGRHOPLAFTDFVNIIG 212  
Db 49 DVPTNPAATILPVHVYPLPQOMKRVASART-----SNLAPGTLDDPIVFDLLNNIGE 103  
Qy 213 DFDAAAGVRCRLPGAATFSGFTLGKLP-RKTLGVKLMKNDEVOAMTYDDGASRRREMOS 271  
Db 104 TFDLQGRFNCVPVGTIVFTFHMCLKAVNPLVYNLMKNEVLVSAYANDGAP-DHETAS 162  
Qy 272 QSVMLALRGGDAVWLLSHDHGYGA-YSNHGKYITFGSGLVYPP 314  
Db 163 NHAILQLFGDQIWLRLH----RGAIYSSWKYSTFSGYLLYOD 202

RESULT 6  
US-10-236-055A-2  
Sequence 2, Application US/10236055A  
Publication No. US20030134328A1  
GENERAL INFORMATION:  
APPLICANT: Basham, Beth E.  
APPLICANT: Foreysche, Ian  
APPLICANT: Gorman, Daniel M.  
APPLICANT: Mattson, Jeanine  
APPLICANT: Moshrefi, Mehrdad  
APPLICANT: Parham, Christi  
TITLE OF INVENTION: MAMMALIAN GENES, RELATED REAGENTS  
FILE REFERENCE: DX01343K  
CURRENT APPLICATION NUMBER: US/10/236,055A  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/317,988  
PRIOR FILING DATE: 2001-09-06  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 2  
LENGTH: 420  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-236-055A-2

Query Match 13.3%; Score 230; DB 12; Length 420;  
Best Local Similarity 38.4%; Pred. No. 3.7e-13;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

Qy 163 DAPARGP-----PAPPEPSAFSAARTSLVSGDAGPGRHOPLAFTDFVNIIG 212  
Db 267 DVPTNPAATILPVHVYPLPQOMKRVASART-----SNLAPGTLDDPIVFDLLNNIGE 321  
Qy 213 DFDAAAGVRCRLPGAATFSGFTLGKLP-RKTLGVKLMKNDEVOAMTYDDGASRRREMOS 271  
Db 322 TFDLQGRFNCVPVGTIVFTFHMCLKAVNPLVYNLMKNEVLVSAYANDGAP-DHETAS 380  
Qy 272 QSVMLALRGGDAVWLLSHDHGYGA-YSNHGKYITFGSGLVYPP 314  
Db 381 NHAILQLFGDQIWLRLH----RGAIYSSWKYSTFSGYLLYOD 420

RESULT 7  
US-09-738-973-185  
Sequence 185, Application US/09738973  
Patent No. US20020110563A1  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Henderson, Robert A.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Flinn, Steven P.  
APPLICANT: Mohamath, Raodoh  
APPLICANT: Algate, Paul A.  
APPLICANT: Secrist, Heather  
APPLICANT: Indriatis, Carol Yoseph  
APPLICANT: Benson, Darin R.  
APPLICANT: Elliot, Mark  
APPLICANT: Mannion, Jane  
APPLICANT: Kalos, Michael D.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

;; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
;; FILE REFERENCE: 210121.475C9  
;; CURRENT APPLICATION NUMBER: US/09/738,973  
;; CURRENT FILING DATE: 2000-12-14  
;; NUMBER OF SEQ ID NOS: 587  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 185  
;; LENGTH: 746  
;; TYPE: PRT  
;; ORGANISM: Homo sapien  
US-09-738-973-185

Query Match 13.3%; Score 230; DB 10; Length 746;  
Best Local Similarity 38.4%; Pred. No. 8e-13;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 163 DAPARGP-----PAPPEPSAFAARTSLVSGSDAGPGRHOPPLAFDTEFVNIG 212  
DB 593 DVPVTNPAATILPPIVHYVPLPQMKRVAFSAART-----SNLAPGTLDPPIVFDLLNNIGE 647  
QY 213 DFDAAAGVFRCLPGAYFFSFTLGKLP-RKTLISYKLMKNRDEVQAMIVDDGASRRREMOS 271  
DB 648 TFDLQGRFNCPPVNGTYVFIFFHMLKLAIVNPLVYVNLKMEBEVLVSAYANDGAP-DHETAS 706  
QY 272 QSWMLALRGDAVWMLSHDHGYGA-YSNHGKYYTFSGFLVYPD 314  
DB 707 NHAILQLFQGDQIWLRLH---RGAIYSSWKYSTFSGYLLYOD 746

RESULT 8  
US-09-854-133-185  
; Sequence 185, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamed, Robert A.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854.133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 185  
; LENGTH: 746  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-09-854-133-185

Query Match 13.3%; Score 230; DB 10; Length 746;  
Best Local Similarity 38.4%; Pred. No. 8e-13;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 163 DAPARGP-----PAPPEPSAFAARTSLVSGSDAGPGRHOPPLAFDTEFVNIG 212  
DB 593 DVPVTNPAATILPPIVHYVPLPQMKRVAFSAART-----SNLAPGTLDPPIVFDLLNNIGE 647  
QY 213 DFDAAAGVFRCLPGAYFFSFTLGKLP-RKTLISYKLMKNRDEVQAMIVDDGASRRREMOS 271  
DB 648 TFDLQGRFNCPPVNGTYVFIFFHMLKLAIVNPLVYVNLKMEBEVLVSAYANDGAP-DHETAS 706  
QY 272 QSWMLALRGDAVWMLSHDHGYGA-YSNHGKYYTFSGFLVYPD 314  
DB 707 NHAILQLFQGDQIWLRLH---RGAIYSSWKYSTFSGYLLYOD 746

RESULT 9  
US-10-144-649A-185  
; Sequence 185, Application US/10144649A  
; Publication No. US20030118599A1

;; GENERAL INFORMATION:  
;; APPLICANT: Lodes, Michael J.  
;; APPLICANT: Wang, Tongtong  
;; APPLICANT: Fan, Liqun  
;; APPLICANT: Algate, Paul A.  
;; APPLICANT: McNeill, Patricia D.  
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
;; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
;; FILE REFERENCE: 210121.475C11  
;; CURRENT APPLICATION NUMBER: US/10/144,649A  
;; CURRENT FILING DATE: 2002-08-21  
;; NUMBER OF SEQ ID NOS: 749  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 185  
;; LENGTH: 746  
;; TYPE: PRT  
;; ORGANISM: Homo sapien  
US-10-144-649A-185

Query Match 13.3%; Score 230; DB 15; Length 746;  
Best Local Similarity 38.4%; Pred. No. 8e-13;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 163 DAPARGP-----PAPPEPSAFAARTSLVSGSDAGPGRHOPPLAFDTEFVNIG 212  
DB 593 DVPVTNPAATILPPIVHYVPLPQMKRVAFSAART-----SNLAPGTLDPPIVFDLLNNIGE 647  
QY 213 DFDAAAGVFRCLPGAYFFSFTLGKLP-RKTLISYKLMKNRDEVQAMIVDDGASRRREMOS 271  
DB 648 TFDLQGRFNCPPVNGTYVFIFFHMLKLAIVNPLVYVNLKMEBEVLVSAYANDGAP-DHETAS 706  
QY 272 QSWMLALRGDAVWMLSHDHGYGA-YSNHGKYYTFSGFLVYPD 314  
DB 707 NHAILQLFQGDQIWLRLH---RGAIYSSWKYSTFSGYLLYOD 746

RESULT 10  
US-10-309-422-10  
; Sequence 10, Application US/10309422  
; Publication No. US20030139587A1  
; GENERAL INFORMATION:  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wiganowski, Nathaniel L.  
; APPLICANT: Turner, C. Alexander Jr.  
; TITLE OF INVENTION: No. US20030139587A1 Human Proteins and Polynucleotides Encodin  
; FILE REFERENCE: LEX-0142-USA  
; CURRENT APPLICATION NUMBER: US/10/309,422  
; CURRENT FILING DATE: 2002-12-03  
; PRIOR APPLICATION NUMBER: US/09/798,771  
; PRIOR FILING DATE: 2001-03-02  
; PRIOR APPLICATION NUMBER: US 60/186,557  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 908  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-309-422-10

Query Match 13.3%; Score 230; DB 12; Length 908;  
Best Local Similarity 38.4%; Pred. No. 1e-12;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 163 DAPARGP-----PAPPEPSAFAARTSLVSGSDAGPGRHOPPLAFDTEFVNIG 212  
DB 755 DVPVTNPAATILPPIVHYVPLPQMKRVAFSAART-----SNLAPGTLDPPIVFDLLNNIGE 809  
QY 213 DFDAAAGVFRCLPGAYFFSFTLGKLP-RKTLISYKLMKNRDEVQAMIVDDGASRRREMOS 271  
DB 810 TFDLQGRFNCPPVNGTYVFIFFHMLKLAIVNPLVYVNLKMEBEVLVSAYANDGAP-DHETAS 866  
QY 272 QSWMLALRGDAVWMLSHDHGYGA-YSNHGKYYTFSGFLVYPD 314

Db 869 NHAILQLFGGQIWLRLH---RGAIVGSSWKSTFSGYLLYQD 908

RESULT 11

US-10-309-422-22

Sequence 22, Application US/10309422

Publication No. US20030139587A1

GENERAL INFORMATION:

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

APPLICANT: Turner, C. Alexander Jr.

TITLE OF INVENTION: No. US20030139587A1el Human Proteins and Polynucleotides Encoding

FILE REFERENCE: LEX-0142-USA

CURRENT APPLICATION NUMBER: US/10/309,422

CURRENT FILING DATE: 2002-12-03

PRIOR APPLICATION NUMBER: US/09/798, 771

PRIOR FILING DATE: 2001-03-02

PRIOR APPLICATION NUMBER: US 60/186,557

PRIOR FILING DATE: 2000-03-02

NUMBER OF SEQ ID NOS: 43

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 22

LENGTH: 909

TYPE: PRT

ORGANISM: homo sapiens

US-10-309-422-22

Query Match 13.3%; Score 230; DB 12; Length 909;

Best Local Similarity 38.4%; Pred. No. 1,le-12;

Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

Qy 163 DAPARGP-----PAPPEPSAFAARTSLVSGDAGPGRHOPLAFDTFEVNIIG 212

Db 756 DVPVTNPATILPVHYVPLPQOMRVAFSAART-----SNLAPGTLDDQPIVFDLLNNLIGE 810

Qy 213 DFDAAAGVRCRLPGAYFFSFTLGKLP-RKTLGVKLMKRDVQAMITDDGASRRREMOS 271

Db 811 TFDLQGRFNCVPVNGTYVTFIHMILKLVNVPPLYVNLKNEEVLVSAYANDGAP-DHETAS 869

Qy 272 QSVMLALRRGDVAWLLSHDHGYGA-YSNHGKYITFSGFLVYPD 314

Db 870 NHAILQLFGGQIWLRLH---RGAIVGSSWKSTFSGYLLYQD 909

RESULT 12

US-10-309-422-14

Sequence 14, Application US/10309422

Publication No. US20030139587A1

GENERAL INFORMATION:

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

APPLICANT: Turner, C. Alexander Jr.

TITLE OF INVENTION: No. US20030139587A1el Human Proteins and Polynucleotides Encoding

FILE REFERENCE: LEX-0142-USA

CURRENT APPLICATION NUMBER: US/10/309,422

CURRENT FILING DATE: 2002-12-03

PRIOR APPLICATION NUMBER: US/09/798, 771

PRIOR FILING DATE: 2001-03-02

PRIOR APPLICATION NUMBER: US 60/186,557

PRIOR FILING DATE: 2000-03-02

NUMBER OF SEQ ID NOS: 43

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 14

LENGTH: 957

TYPE: PRT

ORGANISM: homo sapiens

US-10-309-422-14

Query Match 13.3%; Score 230; DB 12; Length 957;

Best Local Similarity 38.4%; Pred. No. 1,le-12;

Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

Qy 163 DAPARGP-----PAPPEPSAFAARTSLVSGDAGPGRHOPLAFDTFEVNIIG 212

Db 804 DVPVTNPATILPVHYVPLPQOMRVAFSAART-----SNLAPGTLDDQPIVFDLLNNLIGE 858

Qy 213 DFDAAAGVRCRLPGAYFFSFTLGKLP-RKTLGVKLMKRDVQAMITDDGASRRREMOS 271

Db 859 TFDLQGRFNCVPVNGTYVTFIHMILKLVNVPPLYVNLKNEEVLVSAYANDGAP-DHETAS 917

Qy 272 QSVMLALRRGDVAWLLSHDHGYGA-YSNHGKYITFSGFLVYPD 314

Db 918 NHAILQLFGGQIWLRLH---RGAIVGSSWKSTFSGYLLYQD 957

RESULT 13

US-10-309-422-26

Sequence 26, Application US/10309422

Publication No. US20030139587A1

GENERAL INFORMATION:

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

APPLICANT: Turner, C. Alexander Jr.

TITLE OF INVENTION: No. US20030139587A1el Human Proteins and Polynucleotides Encoding

FILE REFERENCE: LEX-0142-USA

CURRENT APPLICATION NUMBER: US/10/309,422

CURRENT FILING DATE: 2002-12-03

PRIOR APPLICATION NUMBER: US/09/798, 771

PRIOR FILING DATE: 2001-03-02

PRIOR APPLICATION NUMBER: US 60/186,557

PRIOR FILING DATE: 2000-03-02

NUMBER OF SEQ ID NOS: 43

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 26

LENGTH: 958

TYPE: PRT

ORGANISM: homo sapiens

US-10-309-422-26

Query Match 13.3%; Score 230; DB 12; Length 958;

Best Local Similarity 38.4%; Pred. No. 1,le-12;

Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

Qy 163 DAPARGP-----PAPPEPSAFAARTSLVSGDAGPGRHOPLAFDTFEVNIIG 212

Db 805 DVPVTNPATILPVHYVPLPQOMRVAFSAART-----SNLAPGTLDDQPIVFDLLNNLIGE 859

Qy 213 DFDAAAGVRCRLPGAYFFSFTLGKLP-RKTLGVKLMKRDVQAMITDDGASRRREMOS 271

Db 860 TFDLQGRFNCVPVNGTYVTFIHMILKLVNVPPLYVNLKNEEVLVSAYANDGAP-DHETAS 918

Qy 272 QSVMLALRRGDVAWLLSHDHGYGA-YSNHGKYITFSGFLVYPD 314

Db 919 NHAILQLFGGQIWLRLH---RGAIVGSSWKSTFSGYLLYQD 958

RESULT 14

US-10-309-422-18

Sequence 18, Application US/10309422

Publication No. US20030139587A1

GENERAL INFORMATION:

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

APPLICANT: Turner, C. Alexander Jr.

TITLE OF INVENTION: No. US20030139587A1el Human Proteins and Polynucleotides Encoding

FILE REFERENCE: LEX-0142-USA

CURRENT APPLICATION NUMBER: US/10/309,422

CURRENT FILING DATE: 2002-12-03

PRIOR APPLICATION NUMBER: US/09/798, 771

PRIOR FILING DATE: 2001-03-02

PRIOR APPLICATION NUMBER: US 60/186,557

PRIOR FILING DATE: 2000-03-02

NUMBER OF SEQ ID NOS: 43

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 18



LENGTH: 992  
TYPE: PRT  
ORGANISM: homo sapiens  
US-10-309-422-18

Query Match 13.3% Score 230; DB 12; Length 992;  
Best Local Similarity 38.4% Pred. No. 1.2e-12;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 163 DAPARGP-----PAPEPRSAFSAARTSLVGSADGPGPRHOPAFDTEFVNIG 212  
DB 839 DVPYTNPAATILPVHVYPLPQOMRVAFSART-----SNLAPGTLDPQIVFDLLNNIG 893  
QY 213 DFDAAGVFRCLPGAFSTLGLP-RKTLGVKLMKNRDEVQAMTYDQASRRREMOS 271  
DB 894 TFDQLGRFNCVPVNGTVYFIFHMLKLVANVPLVYVLMKNBEVLVSAYANDGAP-DHETAS 952  
QY 272 QSVMLALRRGDVWLLSHDHGYGA-YSNHGKXITFSGFLVYPP 314  
DB 953 NHAILQLFGDQIWLRLH---RGAIYGSWKSTFSGYLLYOD 992

RESULT 15

US-10-309-422-30  
Sequence 30, Application US/10309422  
Publication No. US20030139587A1  
GENERAL INFORMATION:  
APPLICANT: Walke, D. Wade  
APPLICANT: Milgowski, Nathaniel L.  
APPLICANT: Turner, C. Alexander Jr.  
TITLE OF INVENTION: No. US20030139587A1 Human Proteins and Polynucleotides Encoding  
FILE REFERENCE: LEX-0142-USA  
CURRENT APPLICATION NUMBER: US/10/309,422  
CURRENT FILING DATE: 2002-12-03  
PRIOR APPLICATION NUMBER: US/09/798,771  
PRIOR FILING DATE: 2001-03-02  
PRIOR APPLICATION NUMBER: US 60/186,557  
PRIOR FILING DATE: 2000-03-02  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 30  
LENGTH: 993  
TYPE: PRT  
ORGANISM: homo sapiens  
US-10-309-422-30

Query Match 13.3% Score 230; DB 12; Length 993;  
Best Local Similarity 38.4% Pred. No. 1.2e-12;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 163 DAPARGP-----PAPEPRSAFSAARTSLVGSADGPGPRHOPAFDTEFVNIG 212  
DB 840 DVPYTNPAATILPVHVYPLPQOMRVAFSART-----SNLAPGTLDPQIVFDLLNNIG 894  
QY 213 DFDAAGVFRCLPGAFSTLGLP-RKTLGVKLMKNRDEVQAMTYDQASRRREMOS 271  
DB 895 TFDQLGRFNCVPVNGTVYFIFHMLKLVANVPLVYVLMKNBEVLVSAYANDGAP-DHETAS 953  
QY 272 QSVMLALRRGDVWLLSHDHGYGA-YSNHGKXITFSGFLVYPP 314  
DB 954 NHAILQLFGDQIWLRLH---RGAIYGSWKSTFSGYLLYOD 993

Search completed: January 12, 2004, 08:33:35  
Job time : 42.0579 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:14 ; Search time 16.7455 Seconds  
(without alignments)  
831.284 Million cell updates/sec

Title: US-10-085-167-2

Perfect score: 1726

Sequence: 1 MLPULLGLGPACWALGPT.....LVFPDLAPAPPGIGASSELL 329

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	230	13.3	746	4	US-09-370-838-185 Sequence 185, App
2	230.5	12.8	281	3	US-09-118-408-44 Sequence 44, Appl
3	220.5	12.8	281	4	US-09-506-855-44 Sequence 44, Appl
4	220.5	12.8	281	4	US-09-911-176B-44 Sequence 44, Appl
5	220.5	12.8	281	4	US-09-619-740-44 Sequence 44, Appl
6	220.5	12.8	281	4	US-09-506-852-44 Sequence 44, Appl
7	215	12.5	281	3	US-09-118-408-2 Sequence 2, Appl
8	215	12.5	281	4	US-09-506-855-2 Sequence 2, Appl
9	215	12.5	281	4	US-09-911-176B-2 Sequence 2, Appl
10	215	12.5	281	4	US-09-619-740-2 Sequence 2, Appl
11	215	12.5	281	4	US-09-506-852-2 Sequence 2, Appl
12	213	12.3	228	4	US-09-336-536-4 Sequence 4, Appl
13	213	12.3	243	3	US-09-140-804-2 Sequence 2, Appl
14	213	12.3	243	4	US-09-336-536-3 Sequence 3, Appl
15	213	12.3	243	4	US-09-686-838B-2 Sequence 2, Appl
16	207	12.0	243	3	US-09-188-930-295 Sequence 295, App
17	207	12.0	243	4	US-09-312-283C-295 Sequence 295, App
18	207	12.0	294	3	US-09-188-930-294 Sequence 294, App
19	207	12.0	294	4	US-09-312-283C-294 Sequence 294, App
20	204	11.8	285	4	US-09-312-283C-382 Sequence 382, App
21	202	11.7	228	4	US-09-336-536-11 Sequence 11, Appl
22	202	11.7	243	4	US-09-336-536-10 Sequence 10, Appl
23	200	11.6	259	4	US-09-996-243-47 Sequence 47, Appl
24	198	11.5	260	4	US-09-489-847-198 Sequence 198, App
25	198	11.5	287	4	US-09-489-847-349 Sequence 349, App
26	196	11.4	231	4	US-09-530-423-2 Sequence 2, Appl
27	196	11.4	244	2	US-08-463-911-7 Sequence 7, Appl

28	196	11.4	244	3	US-09-140-804-3 Sequence 3, Appl
29	196	11.4	244	4	US-09-336-536-20 Sequence 20, Appl
30	196	11.4	244	4	US-09-530-423-1 Sequence 1, Appl
31	196	11.4	244	4	US-09-686-838B-3 Sequence 3, Appl
32	196	11.4	244	4	US-09-911-176B-48 Sequence 48, Appl
33	196	11.4	244	4	US-09-552-225A-3 Sequence 3, Appl
34	196	11.4	244	4	US-09-619-740-51 Sequence 51, Appl
35	196	11.4	244	4	US-09-776-976-6 Sequence 6, Appl
36	196	11.4	244	4	US-09-569-852B-6 Sequence 6, Appl
37	196	11.4	247	2	US-08-463-911-2 Sequence 2, Appl
38	196	11.4	247	3	US-09-140-804-8 Sequence 8, Appl
39	196	11.4	247	3	US-09-118-408-3 Sequence 3, Appl
40	196	11.4	247	4	US-09-506-855-3 Sequence 3, Appl
41	196	11.4	247	4	US-09-686-838B-8 Sequence 8, Appl
42	196	11.4	247	4	US-09-911-176B-3 Sequence 3, Appl
43	196	11.4	247	4	US-09-619-740-3 Sequence 3, Appl
44	196	11.4	247	4	US-09-776-976-4 Sequence 4, Appl
45	196	11.4	247	4	US-09-776-976-4 Sequence 4, Appl

#### ALIGNMENTS

```
RESULT 1
US-09-370-838-185
; Sequence 185, Application US/09370838
; Patent No. 644425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Monamath, Rosdoh
; TITLE OF INVENTION: SECRETIST, HEALER
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370.838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285.323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 185
; LENGTH: 746
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-370-838-185

Query Match      13.3%; Score 230; DB 4; Length 746;
Best Local Similarity 38.4%; Pred. No. 9.1e-15;
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 163 DAPARGP-----PAPPEPRSAFSAARTSLVSGDAGPGRHQPLAFDFEFVIGG 212
DB 593 DVEVTNPAATILPVHYVPLPQOKRVAFSAART-----SNLAPGTLDPVFDLLNLNGE 647
QY 213 DFDAAAGVFCRCPLGAVFFSFTLGKLP-RKTLSEVKLMKNRDEVOAMIVDDGASRRRMOG 271
DB 648 TFLQIGRFPCPNNGTVVFIHMLKLVNPLVYVNLKREVLVSAYANDGAP-DHETAS 706
QY 272 QSMVLAIRGDVAWLISHDHGYGA-YSNKGKITYTSGGLVYPD 314
DB 707 NHAIIQIFQGDQIWLRLH-----RGATVGSSEWKTSTFSGVLYYOD 746

RESULT 2
US-09-118-408-44
; Sequence 44, Application US/09118408A
; Patent No. 6265544
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-30
; CURRENT APPLICATION NUMBER: US/09/118.408A
```

```
; CURRENT FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: 60/053.154
; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 44
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-118-408-44

Query Match      12.8%; Score 220.5; DB 3; Length 281;
Best Local Similarity 39.2%; Pred. No. 2.1e-14;
Matches 56; Conservative 20; Mismatches 58; Indels 9; Gaps 4;

Qy 177 SAFSAARTSLVGSADGPGPRHQPAPTEPFVNIIGDPAAGVRCRLPGAYFFSFTLG 236
Db 146 AAFSVGRKKALHSND-----YFQPVFDTTEFVNLKHFNMFTGKFCYCVPGIYFFSLNVH 200
Qy 237 KLPRKTLGVKLMKNRDEVQAMTYDDGASRRREMOSQVMTALRRGDVAV--LLSHDHGX 294
Db 201 TWMQKETYLIHMKNEEV-VILYAQ-VSDRSIMOSQSLMELREDEVWVRLFKGERENA 258
Qy 295 GAYSNHGKTYTFSGFLVYPDLAP 317
Db 259 IFSDEPDTYITFSGYLVKRPASEP 281

RESULT 3
US-09-506-855-44
; Sequence 44, Application US/09506855
; Patent No. 6448221
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Laesser, Gerald W.
; TITLE OF INVENTION: INHIBITORS FOR USE IN HEMOSTASIS AND
; FILE REFERENCE: 99-12
; CURRENT APPLICATION NUMBER: US/09/506,855
; CURRENT FILING DATE: 2000-02-17
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 44
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-506-855-44

Query Match      12.8%; Score 220.5; DB 4; Length 281;
Best Local Similarity 39.2%; Pred. No. 2.1e-14;
Matches 56; Conservative 20; Mismatches 58; Indels 9; Gaps 4;

Qy 177 SAFSAARTSLVGSADGPGPRHQPAPTEPFVNIIGDPAAGVRCRLPGAYFFSFTLG 236
Db 146 AAFSVGRKKALHSND-----YFQPVFDTTEFVNLKHFNMFTGKFCYCVPGIYFFSLNVH 200
Qy 237 KLPRKTLGVKLMKNRDEVQAMTYDDGASRRREMOSQVMTALRRGDVAV--LLSHDHGX 294
Db 201 TWMQKETYLIHMKNEEV-VILYAQ-VSDRSIMOSQSLMELREDEVWVRLFKGERENA 258
Qy 295 GAYSNHGKTYTFSGFLVYPDLAP 317
Db 259 IFSDEPDTYITFSGYLVKRPASEP 281

RESULT 4
US-09-911-176B-44
; Sequence 44, Application US/09911176B
; Patent No. 6518403
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: ANTIBODIES THAT BIND AN
```

```
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGY
; FILE REFERENCE: 97-30D1
; CURRENT APPLICATION NUMBER: US/09/911,176B
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/118,408
; PRIOR FILING DATE: 1998-07-17
; PRIOR APPLICATION NUMBER: 60/053,154
; PRIOR FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 44
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-911-176B-44

Query Match      12.8%; Score 220.5; DB 4; Length 281;
Best Local Similarity 39.2%; Pred. No. 2.1e-14;
Matches 56; Conservative 20; Mismatches 58; Indels 9; Gaps 4;

Qy 177 SAFSAARTSLVGSADGPGPRHQPAPTEPFVNIIGDPAAGVRCRLPGAYFFSFTLG 236
Db 146 AAFSVGRKKALHSND-----YFQPVFDTTEFVNLKHFNMFTGKFCYCVPGIYFFSLNVH 200
Qy 237 KLPRKTLGVKLMKNRDEVQAMTYDDGASRRREMOSQVMTALRRGDVAV--LLSHDHGX 294
Db 201 TWMQKETYLIHMKNEEV-VILYAQ-VSDRSIMOSQSLMELREDEVWVRLFKGERENA 258
Qy 295 GAYSNHGKTYTFSGFLVYPDLAP 317
Db 259 IFSDEPDTYITFSGYLVKRPASEP 281

RESULT 5
US-09-619-740-44
; Sequence 44, Application US/09619740
; Patent No. 6544946
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Laesser, Gerald W.
; TITLE OF INVENTION: INHIBITORS FOR USE IN HEMOSTASIS AND IMMUNE FUNCTION
; FILE REFERENCE: 99-12C3
; CURRENT APPLICATION NUMBER: US/09/619,740
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/253,604
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/444,794
; PRIOR FILING DATE: 1999-11-22
; PRIOR APPLICATION NUMBER: 09/506,855
; PRIOR FILING DATE: 2000-02-17
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 44
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-619-740-44

Query Match      12.8%; Score 220.5; DB 4; Length 281;
Best Local Similarity 39.2%; Pred. No. 2.1e-14;
Matches 56; Conservative 20; Mismatches 58; Indels 9; Gaps 4;

Qy 177 SAFSAARTSLVGSADGPGPRHQPAPTEPFVNIIGDPAAGVRCRLPGAYFFSFTLG 236
Db 146 AAFSVGRKKALHSND-----YFQPVFDTTEFVNLKHFNMFTGKFCYCVPGIYFFSLNVH 200
Qy 237 KLPRKTLGVKLMKNRDEVQAMTYDDGASRRREMOSQVMTALRRGDVAV--LLSHDHGX 294
Db 201 TWMQKETYLIHMKNEEV-VILYAQ-VSDRSIMOSQSLMELREDEVWVRLFKGERENA 258
Qy 295 GAYSNHGKTYTFSGFLVYPDLAP 317
```



; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 281  
; TYPE: PRF  
; ORGANISM: Homo sapien  
US-09-911-176B-2

Query Match 12.5%; Score 215; DB 4; Length 281;  
Best Local Similarity 30.2%; Pred. No. 7,6e-14;  
Matches 77; Conservative 26; Mismatches 86; Indels 66; Gaps 11;

QY 69 RCRVGAFFSFTACKAPKSLVNLVNRDEVQALAFDEORRPGARPAASQSMLOLDY 128  
DB 75 RCDPDTSMYPATA--VPQINTITLKGEKDR-----GDRG-----LQKXY 113  
QY 129 GDT--VWLRHGAP---HYALGAPGATSGYLIVADADADAPARPPAPPEPRSAFSAAR 183  
DB 114 GKTGSAGARGHTGPGQKSGMGAPEKCKSH--YA-----AFSVGR 152  
QY 184 TRSLVGSADAGPGRH-----QPLAFDTEFVNIGDPPDAAGVFRCLPGAYFSPFTLGKL 238  
DB 153 KK-----PMSNHYYQTVIFDTEFVNLVYDHFMFTGKFCYVPGLYFSLVHTW 202  
QY 239 PRKTLVYKLMKNRDEVQAMLYDDGASRRRMOQSVMALRRGDAVWLLSHDHGCGAYS 298  
DB 203 NQKETYLIHMKNEEVILFAQVG--DRSIMQSGLMELREDOQVWVRLYKGERENAIF 260  
QY 299 NH--GKTYTFSGLV 311  
DB 261 SEEDTYITFSGLV 275

## RESULT 10

US-09-619-740-2  
; Sequence 2, Application US/09619740  
; Patent No. 6544946  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; APPLICANT: Laseer, Gerald W.  
; APPLICANT: Bishop, Paul D.  
; TITLE OF INVENTION: INHIBITORS FOR USE IN HEMOSTASIS AND IMMUNE FUNCTION  
; FILE REFERENCE: 99-12C3  
; CURRENT FILING DATE: US/09/619,740  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/253,604  
; PRIOR FILING DATE: 1999-02-19  
; PRIOR APPLICATION NUMBER: 09/444,794  
; PRIOR FILING DATE: 1999-11-22  
; PRIOR APPLICATION NUMBER: 09/506,855  
; PRIOR FILING DATE: 2000-02-17  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 281  
; TYPE: PRF  
; ORGANISM: Homo sapien  
US-09-619-740-2

Query Match 12.5%; Score 215; DB 4; Length 281;  
Best Local Similarity 30.2%; Pred. No. 7,6e-14;  
Matches 77; Conservative 26; Mismatches 86; Indels 66; Gaps 11;  
QY 69 RCRVGAFFSFTACKAPKSLVNLVNRDEVQALAFDEORRPGARPAASQSMLOLDY 128  
DB 75 RCDPDTSMYPATA--VPQINTITLKGEKDR-----GDRG-----LQKXY 113  
QY 129 GDT--VWLRHGAP---HYALGAPGATSGYLIVADADADAPARPPAPPEPRSAFSAAR 183  
DB 114 GKTGSAGARGHTGPGQKSGMGAPEKCKSH--YA-----AFSVGR 152  
QY 184 TRSLVGSADAGPGRH-----QPLAFDTEFVNIGDPPDAAGVFRCLPGAYFSPFTLGKL 238

DB 153 KK-----PMSNHYYQTVIFDTEFVNLVYDHFMFTGKFCYVPGLYFSLVHTW 202  
QY 239 PRKTLVYKLMKNRDEVQAMLYDDGASRRRMOQSVMALRRGDAVWLLSHDHGCGAYS 298  
DB 203 NQKETYLIHMKNEEVILFAQVG--DRSIMQSGLMELREDOQVWVRLYKGERENAIF 260  
QY 299 NH--GKTYTFSGLV 311  
DB 261 SEEDTYITFSGLV 275

## RESULT 11

US-09-506-852-2  
; Sequence 2, Application US/09506852  
; Patent No. 6566499  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS  
; FILE REFERENCE: 97-30  
; CURRENT FILING DATE: US/09/506,852  
; CURRENT FILING DATE: 2000-02-17  
; EARLIER APPLICATION NUMBER: 60/053,154  
; EARLIER FILING DATE: 1997-07-18  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 281  
; TYPE: PRF  
; ORGANISM: Homo sapien  
US-09-506-852-2

Query Match 12.5%; Score 215; DB 4; Length 281;  
Best Local Similarity 30.2%; Pred. No. 7,6e-14;  
Matches 77; Conservative 26; Mismatches 86; Indels 66; Gaps 11;

QY 69 RCRVGAFFSFTACKAPKSLVNLVNRDEVQALAFDEORRPGARPAASQSMLOLDY 128  
DB 75 RCDPDTSMYPATA--VPQINTITLKGEKDR-----GDRG-----LQKXY 113  
QY 129 GDT--VWLRHGAP---HYALGAPGATSGYLIVADADADAPARPPAPPEPRSAFSAAR 183  
DB 114 GKTGSAGARGHTGPGQKSGMGAPEKCKSH--YA-----AFSVGR 152  
QY 184 TRSLVGSADAGPGRH-----QPLAFDTEFVNIGDPPDAAGVFRCLPGAYFSPFTLGKL 238  
DB 153 KK-----PMSNHYYQTVIFDTEFVNLVYDHFMFTGKFCYVPGLYFSLVHTW 202  
QY 239 PRKTLVYKLMKNRDEVQAMLYDDGASRRRMOQSVMALRRGDAVWLLSHDHGCGAYS 298  
DB 203 NQKETYLIHMKNEEVILFAQVG--DRSIMQSGLMELREDOQVWVRLYKGERENAIF 260  
QY 299 NH--GKTYTFSGLV 311  
DB 261 SEEDTYITFSGLV 275

## RESULT 12

US-09-336-536-4  
; Sequence 4, Application US/09336536  
; Patent No. 6406884  
; GENERAL INFORMATION:  
; APPLICANT: Leidy, K.  
; APPLICANT: McKay, K.  
; APPLICANT: Boscone, S.  
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF  
; FILE REFERENCE: 7853-144  
; CURRENT FILING DATE: US/09/336,536  
; CURRENT FILING DATE: 1999-06-18  
; NUMBER OF SEQ ID NOS: 75  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 228  
; TYPE: PRF

```

; ORGANISM: Homo sapiens
US-09-336-536-4

Query Match      12.3%; Score 213; DB 4; Length 228;
Best Local Similarity 34.9%; Pred. No. 9e-14;
Matches 66; Conservative 18; Mismatches 83; Indels 22; Gaps 6;

QY 138 GAPHYALGAPGATPSG---YLIVYADAD---ADAPARGPAPPP-----EPRSASFAARTR 185
DB 42 GAP-----GAPGKEGEGGRPGLPGRGDPGRGEGAPGPTGPAGECSVPPRSASFAKRSK 97
QY 186 SLVSGDAGPGPRHOPLAFDTEFVNIGSDPDAAAGVFRCLPGAYFSEFTLGLPRKTLISV 245
DB 98 SRV-----PPPSDAPLPFDRLVNVNQGHYDAVTGKFTQVPGVYFFA-VHATVYRASLOF 151
QY 246 KLKMKRDEVQAMITYDDGASRRREMOSQSVMLALRRGDAVWLLSHDHGCGAYSNHGKYIT 305
DB 152 DLVKNGBESIASFPQFGGMPKRPASLSGAMVRLPEPDQVWVQVGVGDYIGIYASIKTDST 211
QY 306 FSGFLVYPD 314
DB 212 FSGFLVYSD 220

RESULT 13
US-09-140-804-2
; Sequence 2, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT APPLICATION NUMBER: US/09/140,804
; CURRENT FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: 60/056,983
; EARLIER FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-140-804-2

Query Match      12.3%; Score 213; DB 3; Length 243;
Best Local Similarity 34.9%; Pred. No. 9.9e-14;
Matches 66; Conservative 18; Mismatches 83; Indels 22; Gaps 6;

QY 138 GAPHYALGAPGATPSG---YLIVYADAD---ADAPARGPAPPP-----EPRSASFAARTR 185
DB 57 GAP-----GAPGKEGEGGRPGLPGRGDPGRGEGAPGPTGPAGECSVPPRSASFAKRSK 112
QY 186 SLVSGDAGPGPRHOPLAFDTEFVNIGSDPDAAAGVFRCLPGAYFSEFTLGLPRKTLISV 245
DB 113 SRV-----PPPSDAPLPFDRLVNVNQGHYDAVTGKFTQVPGVYFFA-VHATVYRASLOF 166
QY 246 KLKMKRDEVQAMITYDDGASRRREMOSQSVMLALRRGDAVWLLSHDHGCGAYSNHGKYIT 305
DB 167 DLVKNGBESIASFPQFGGMPKRPASLSGAMVRLPEPDQVWVQVGVGDYIGIYASIKTDST 226
QY 306 FSGFLVYPD 314
DB 227 FSGFLVYSD 235

RESULT 14
US-09-336-536-3
; Sequence 3, Application US/09336536
; Patent No. 6406884
; GENERAL INFORMATION:
; APPLICANT: Leiby, K.
; APPLICANT: McKay, C.
```

```

; APPLICANT: Bessone, S.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-144
; CURRENT APPLICATION NUMBER: US/09/336,536
; CURRENT FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-336-536-3

Query Match      12.3%; Score 213; DB 4; Length 243;
Best Local Similarity 34.9%; Pred. No. 9.9e-14;
Matches 66; Conservative 18; Mismatches 83; Indels 22; Gaps 6;

QY 138 GAPHYALGAPGATPSG---YLIVYADAD---ADAPARGPAPPP-----EPRSASFAARTR 185
DB 57 GAP-----GAPGKEGEGGRPGLPGRGDPGRGEGAPGPTGPAGECSVPPRSASFAKRSK 112
QY 186 SLVSGDAGPGPRHOPLAFDTEFVNIGSDPDAAAGVFRCLPGAYFSEFTLGLPRKTLISV 245
DB 113 SRV-----PPPSDAPLPFDRLVNVNQGHYDAVTGKFTQVPGVYFFA-VHATVYRASLOF 166
QY 246 KLKMKRDEVQAMITYDDGASRRREMOSQSVMLALRRGDAVWLLSHDHGCGAYSNHGKYIT 305
DB 167 DLVKNGBESIASFPQFGGMPKRPASLSGAMVRLPEPDQVWVQVGVGDYIGIYASIKTDST 226
QY 306 FSGFLVYPD 314
DB 227 FSGFLVYSD 235

RESULT 15
US-09-686-838B-2
; Sequence 2, Application US/09686838B
; Patent No. 6482612
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: Adipocyte-Specific Protein Homologs
; FILE REFERENCE: 97-49D1
; CURRENT APPLICATION NUMBER: US/09/686,838B
; CURRENT FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/140,804
; PRIOR FILING DATE: 1998-08-26
; PRIOR APPLICATION NUMBER: US 60/056,983
; PRIOR FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-686-838B-2

Query Match      12.3%; Score 213; DB 4; Length 243;
Best Local Similarity 34.9%; Pred. No. 9.9e-14;
Matches 66; Conservative 18; Mismatches 83; Indels 22; Gaps 6;

QY 138 GAPHYALGAPGATPSG---YLIVYADAD---ADAPARGPAPPP-----EPRSASFAARTR 185
DB 57 GAP-----GAPGKEGEGGRPGLPGRGDPGRGEGAPGPTGPAGECSVPPRSASFAKRSK 112
QY 186 SLVSGDAGPGPRHOPLAFDTEFVNIGSDPDAAAGVFRCLPGAYFSEFTLGLPRKTLISV 245
DB 113 SRV-----PPPSDAPLPFDRLVNVNQGHYDAVTGKFTQVPGVYFFA-VHATVYRASLOF 166
QY 246 KLKMKRDEVQAMITYDDGASRRREMOSQSVMLALRRGDAVWLLSHDHGCGAYSNHGKYIT 305
DB 167 DLVKNGBESIASFPQFGGMPKRPASLSGAMVRLPEPDQVWVQVGVGDYIGIYASIKTDST 226
```

Oy	306	FSGFLVYPD	314
Db	227	FSGFLVYSD	235

Search completed: January 12, 2004, 08:20:03  
Job time : 17.7455 secs

---



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:09 ; Search time 0.926148 Seconds  
(without alignments)  
1661.397 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_1\_16  
Perfect score: 87  
Sequence: 1 MLPILGLGIPACMA 16

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76:\*

1: PIR1:.\*  
2: PIR2:.\*  
3: PIR3:.\*  
4: PIR4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57	65.5	993	2	148653 mouse developmenta
2	52	59.8	115	2	S72634 type 4 fibrial bl
3	49	56.3	1180	2	T31066 vascular cadherin-
4	46	52.9	283	2	T16945 hypothetical prote
5	45	51.7	303	2	AB3201 conserved hypothet
6	45	51.7	315	2	E75382 hypothetical prote
7	45	51.7	416	1	S06763 calreticulin precu
8	45	51.7	416	2	JH0819 calreticulin precu
9	45	51.7	418	1	A34154 calreticulin precu
10	45	51.7	628	2	D97073 PTS system, beta-g
11	44	50.6	318	2	A55429 11-cis retinol den
12	44	50.6	553	2	F75497 probable transpo
13	44	50.6	1220	2	T18291 patched protein -
14	44	49.4	119	2	C81919 Probable membrane
15	43	49.4	119	2	F81182 conserved hypothet
16	43	49.4	146	2	B75021 hypothetical prote
17	43	49.4	159	2	G75555 conserved hypothet
18	43	49.4	273	1	C70127 guia protein homol
19	43	49.4	275	2	H83325 hypothetical prote
20	43	49.4	341	1	VVVPK2 coat protein VP2 -
21	43	49.4	344	1	VVVPJ5 coat protein VP2 -
22	43	49.4	351	1	VVVPJ2 coat protein VP2 -
23	43	49.4	351	1	VVVPAS coat protein VP2 -
24	43	49.4	352	1	VVVPJ4 coat protein VP2 -
25	43	49.4	368	2	S75652 ABC-type transpor
26	43	49.4	461	2	T11829 ABC2 dehydrogenas
27	43	49.4	558	2	AG2626 ABC transporter, m
28	43	49.4	558	2	F97408 hypothetical prote
29	42.5	48.9	319	1	VVVPJ2 coat protein VP2 -

30	42.5	48.9	319	1	VVVPJ2	coat protein VP2 -
31	42	48.3	227	2	J01997	hypothetical 26k p
32	42	48.3	361	2	AE1979	ABC transporter At
33	42	48.3	403	2	T34923	hypothetical ROK E
34	42	48.3	657	2	D71351	probable primosoma
35	41.5	47.7	265	2	B96913	ABC-type probable
36	41	47.1	66	2	AB3575	hypothetical prote
37	41	47.1	110	2	S22893	T-cell receptor al
38	41	47.1	121	2	H81892	hypothetical prote
39	41	47.1	229	2	T18629	hypothetical prote
40	41	47.1	265	2	A82000	phosphatidate cyti
41	41	47.1	265	2	B81229	phosphatidate cyti
42	41	47.1	267	2	T36675	probable integral
43	41	47.1	277	2	I17552	Ox40 homolog - hum
44	41	47.1	341	2	A69209	conserved hypothet
45	41	47.1	345	1	VVVPJ2	coat protein VP2 -

## ALIGNMENTS

RESULT 1  
148653  
mouse developmental kinase 5 (MDK5) - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 18-Jun-1999  
C:Accession: 148653  
R:Ciossek, T.; Lerch, M.M.; Ullrich, A.  
Oncogene 11, 2085-2095, 1995  
A>Title: Cloning, characterization, and differential expression of MDK2 and MDK5, two nov  
A:Reference number: 148652; MUID:96074837; PMID:7478528  
A:Accession: 148653  
A>Status: Preliminary; translated from GB/EMBL/DBJ  
A:Residues: 1-993 <RES>  
A:Molecule type: mRNA  
A:Cross-references: EMBL:249086; NID:G1089899; PIDN:CAA88910.1; PID:G1089900  
C:Superfamily: protein-tyrosine kinase, receptor type eph; fibronectin type III repeat hc  
C:Keywords: ATP; transmembrane protein  
F:626-894/Domain: protein kinase homology <KIN>  
F:634-642/Region: protein kinase ATP-binding motif  
F:917-983/Domain: SAM homology <SAM>

Query Match 65.5%; Score 57; DB 2; Length 993;  
Best Local Similarity 68.8%; Pred. No. 0.73;  
Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
OY 1 MLPILGLGIPACMA 16  
DB 14 LAPILLPLIPACMA 29

RESULT 2  
S72634  
type 4 fibrial biogenesis protein piliy2 PA455 [imported] - Pseudomonas aeruginosa  
C:Species: Pseudomonas aeruginosa  
C>Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 31-Dec-2000  
C:Accession: S72634; E83076  
F:Alm, R.A.; Hallinan, J.P.; Watson, A.A.; Mattick, J.S.  
Mol. Microbiol. 22, 161-173, 1996  
A>Title: Fibrial biogenesis genes of Pseudomonas aeruginosa: pilW and pilX increase the  
u.  
A:Reference number: S72631; MUID:97055431; PMID:8899718  
A:Accession: S72634  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-115 <ALM>  
A:Cross-references: EMBL:L76605; NID:G1246298; PIDN:AAA93503.1; PID:G1246303  
A:Experimental source: strain P401  
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, April 1996  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bri  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidig, K.; Lim,  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000

A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho  
 A:Reference number: AB2950; MUID:20437377; PMID:10984043  
 A:Accession: E83076  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-115 <STO>  
 A:Cross-references: GB:AE004869; GB:AE004091; NID:g9950793; PIDN:AA607943.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: piliY2; PA4555  
 C:Superfamily: *Pseudomonas aeruginosa* piliY2 protein  
 C:Keywords: fimbria

Query Match 59.8%; Score 52; DB 2; Length 115;  
 Best Local Similarity 56.2%; Pred. No. 0.6;  
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 MFLPLGLGPAACWA 16  
 :|||:|  
 Db 3 VLPMLALAVPGLCWA 18

RESULT 3  
 T31066  
 A:Title: vascular cadherin-2 - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 21-Jul-2000  
 C:Accession: T31066  
 R:Tejo, P.; Breviario, F.; Huber, P.; Panzerl, C.; Dejana, E.  
 J. Biol. Chem. 273, 17565-17572, 1998  
 A:Title: Identification of a novel cadherin (vascular endothelial cadherin-2) located at  
 A:Reference number: Z20962; MUID:98316322; PMID:9651350  
 A:Accession: T31066  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1180 <TEB>  
 A:Cross-references: EMBL:X08715; NID:g2764993; PIDN:CAA69965.1; PID:g2764994  
 A:Experimental source: brain capillary

Query Match 56.3%; Score 49; DB 2; Length 1180;  
 Best Local Similarity 69.2%; Pred. No. 14;  
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 MFLPLGLGPAACWA 13  
 :|||:|  
 Db 4 LVPPLGLGPGS 16

RESULT 4  
 T16945  
 A:Title: hypothetical protein T27F7.3 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C:Accession: T16945  
 R:Hallsworth, K.  
 Submitted to the EMBL Data Library, May 1996  
 A:Description: The sequence of C. elegans cosmid T27F7.  
 A:Reference number: Z18613  
 A:Accession: T16945  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-283 <HAU>  
 A:Cross-references: EMBL:U58762; NID:g1330398; PID:g1330400; PIDN:AA800719.1; GSPDB:GN00  
 A:Experimental source: strain Bristol N2; clone T27F7  
 C:Genetics:  
 A:Gene: CESP:T27F7.3  
 A:Map position: 2  
 A:introns: 11/1, 45/1, 63/1, 130/3, 248/3

Query Match 52.9%; Score 46; DB 2; Length 283;  
 Best Local Similarity 63.6%; Pred. No. 11;  
 Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 MFLPLGLGPAACWA 11  
 :|||:|  
 Db 162 MFLPLGLGPAACWA 172

RESULT 5  
 AB3201  
 A:Title: conserved hypothetical protein Atu5339 (imported) - *Agrobacterium tumefaciens* (strain C58  
 C:Species: *Agrobacterium tumefaciens*  
 C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
 C:Accession: AB3201  
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; McChelle  
 ; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, I  
 ster, E.W.  
 A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.  
 A:Reference number: AB2577; MUID:21608550; PMID:11743193  
 A:Accession: AB3201  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-303 <KUR>  
 A:Cross-references: GB:AE008687; PIDN:AAL46027.1; PID:g17743784; GSPDB:GN00188  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: Atu5339  
 A:Genome: plasmid

Query Match 51.7%; Score 45; DB 2; Length 303;  
 Best Local Similarity 81.8%; Pred. No. 17;  
 Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 LILGLGPAAC 14  
 :|||:|  
 Db 70 ILGLGPAAC 80

RESULT 6  
 E75382  
 A:Title: hypothetical protein - *Deinococcus radiodurans* (strain R1)  
 C:Species: *Deinococcus radiodurans*  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 28-Jul-2000  
 C:Accession: E75382  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; I  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.  
 A:Reference number: A75250; MUID:20036896; PMID:10567266  
 A:Accession: E75382  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-315 <WHI>  
 A:Cross-references: GB:AE001998; GB:AE000513; NID:g6459302; PIDN:AA11113.1; PID:g6459312  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR1541  
 A:Map position: 1  
 C:Superfamily: *Deinococcus radiodurans* hypothetical protein DR1541

Query Match 51.7%; Score 45; DB 2; Length 315;  
 Best Local Similarity 53.3%; Pred. No. 18;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 MFLPLGLGPAACWA 15  
 :|||:|  
 Db 26 LVPPLGLGPAACWA 40

RESULT 7  
 S06763  
 A:Title: calreticulin precursor - mouse

N:Alternate names: 55K calcium-binding reticuloplasmin; calregulin  
C:Species: Mus musculus (house mouse)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: S06763; JCI1444; PC1233; A57498  
R:Smith, M.J.; Koch, G.L.E.  
EMBO J. 8, 3581-3586, 1989  
A:Title: Multiple zones in the sequence of calreticulin (CPRP5, calregulin, HACBP), a me  
A:Reference number: S06763; MUID:90059955; PMID:2583110  
A:Accession: S06763  
A:Molecule type: DNA  
A:Residues: 1-416 <SM>  
A:Cross-references: EMBL:X14926; NID:950567; PIDN:CAA33053.1; PID:950568  
R:Mazzarella, R.A.; Gold, P.; Cunningham, M.; Green, M.  
Gene 120, 217-225, 1992  
A:Title: Determination of the sequence of an expressible cDNA clone encoding ERp60/calre  
A:Reference number: JCI1444; MUID:9301037; PMID:1398135  
A:Accession: JCI1444  
A:Molecule type: mRNA  
A:Residues: 1-416 <MA2>  
A:Cross-references: GB:M92988; NID:9193084; PIDN:AAA37569.1; PID:9193085  
A:Accession: PC1233  
A:Molecule type: protein  
A:Residues: 18-41 <MA2>  
R:White, T.K.; Zhu, O.; Tanzer, M.L.  
J. Biol. Chem. 270, 15926-15929, 1995  
A:Title: Cell surface calreticulin is a putative mannoside lectin which triggers mouse m  
A:Reference number: A57498; MUID:95332280; PMID:7608143  
A:Accession: A57498  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 74-80;142-151;166-193 <WH1>  
C:Superfamily: calreticulin  
C:Keywords: calcium binding  
F:1-17/Domain: signal sequence #status predicted <SIG>  
F:18-416/Product: calregulin #status experimental <MAT>  
F:413-416/Region: endoplasmic reticulum retention signal

Query Match 51.7%; Score 45; DB 1; Length 416;  
Best Local Similarity 83.3%; Pred. No. 23;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LPLLLGLGPPAA 13  
Db 5 VPLLLGLGLAA 16

RESULT 8  
JH0819  
N:Alternate names: precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 30-Sep-1993 #sequence\_revision 20-Aug-1994 #text\_change 20-Jun-2000  
C:Accession: JH0819; A49176; S11205; PC1109; S45036; S39372; A34473; S13045  
R:Nakamura, M.; Moriya, M.; Baba, T.; Michikawa, Y.; Yamanobe, T.; Arai, K.; Okinaga, S.  
Exp. Cell Res. 205, 101-110, 1993  
A:Title: An endoplasmic reticulum protein, calreticulin, is transported into the acrosom  
A:Reference number: A49176; MUID:93202172; PMID:8453984  
A:Accession: JH0819  
A:Molecule type: mRNA  
A:Residues: 1-416 <NMK>  
A:Cross-references: GB:D78308; NID:91089798; PIDN:BA11345.1; PID:91845572  
A:Accession: A49176  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-416 <NA2>  
A:Cross-references: GB:D78308; NID:91089798; PIDN:BA11345.1; PID:91845572  
A:Experimental source: Sprague-Dawley, spermatogenic cells  
A>Note: sequence extracted from NCBI Backbone (NCBI:127639, NCBI:127643)  
R:Mutthy, K.K.; Banville, D.; Srikant, C.B.; Carrier, F.; Holmes, C.; Bell, A.; Patel, Y.  
Nucleic Acids Res. 18, 4933, 1990  
A:Title: Structural homology between the rat calreticulin gene product and the Onchocerc  
A:Reference number: S11205; MUID:90370496; PMID:2395661  
A:Accession: S11205

A:Molecule type: mRNA  
A:Residues: 1-416 <MR>  
A:Cross-references: EMBL:X53363; NID:955854; PIDN:CAA37446.1; PID:955855  
R:Nakamura, M.; Michikawa, Y.; Baba, T.; Okinaga, S.; Arai, K.  
Biochem. Biophys. Res. Commun. 186, 668-673, 1992  
A:Title: Calreticulin is present in the acrosome of spermatozoa of rat testis.  
A:Reference number: PC1109; MUID:92360010; PMID:1497655  
A:Accession: PC1109  
A:Molecule type: protein  
A:Residues: 18-32 <NAK2>  
A:Experimental source: testis, strain Sprague-Dawley  
R:Seemitschen, B.; Fellekrug, J.; van Nguyen, P.; Diekmann, W.; Robinson, D.G.; Mieskes,  
submitted to the EMBL Data Library, May 1994  
A:Description: Retention and retrieval: both mechanisms cooperate to maintain calreticuli  
A:Reference number: S45036  
A:Accession: S45036  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-416 <SOE>  
A:Cross-references: EMBL:X79327; NID:9488840; PIDN:CAA55890.1; PID:9488841  
R:Lione, Y.C.; Bailly, A.; Latruffe, N.  
submitted to the EMBL Data Library, December 1998  
A:Reference number: S04867  
A:Accession: S04867  
A:Molecule type: mRNA  
A:Residues: R, 270-358, 'MAG' <LON>  
A:Cross-references: EMBL:X13702; NID:956055; PIDN:CAA31987.1; PID:930260  
A>Note: the authors designated the protein as D-beta-hydroxybutyrate dehydrogenase  
R:Yokoi, T.; Nagayama, S.; Kajiwara, R.; Kawaguchi, Y.; Horuchi, R.; Kametaki, T.  
Biochim. Biophys. Acta 1158, 339-344, 1993  
A:Title: Identification of protein disulfide isomerase and calreticulin as autoimmune ant  
A:Reference number: S39371; MUID:94072621; PMID:8251535  
A:Accession: S39372  
A:Molecule type: protein  
A:Residues: 18-23, 'X', 25-32 <YOK>  
R:Van, P.N.; Peter, F.; Soelling, H.D.  
J. Biol. Chem. 264, 17494-17501, 1989  
A:Title: Four intracisternal calcium-binding glycoproteins from rat liver microsomes wit  
A:Reference number: A34473; MUID:90089920; PMID:2793869  
A:Accession: A34473  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 18-36 <VAN>  
R:Teves, S.; de Mattei, M.; Lanfredi, M.; Villa, A.; Green, N.M.; MacLennan, D.H.; Meld  
Biochem. J. 271, 473-480, 1990  
A:Title: Calreticulin is a candidate for a calsequestrin-like function in Ca(2+)-storage  
A:Reference number: S13045; MUID:91054414; PMID:2241926  
A:Accession: S13045  
A:Molecule type: protein  
A:Residues: 18-29 <TRB>  
C:Superfamily: calreticulin  
C:Keywords: calcium binding; glycoprotein  
F:1-17/Domain: signal sequence #status predicted <SIG>  
F:18-416/Product: calreticulin #status experimental <MAT>  
F:204-212/Region: nuclear location signal  
F:413-416/Region: endoplasmic reticulum retention signal  
F:344/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 51.7%; Score 45; DB 2; Length 416;  
Best Local Similarity 83.3%; Pred. No. 23;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LPLLLGLGPPAA 13  
Db 5 VPLLLGLGLAA 16

RESULT 9  
A34154  
N:Alternate names: precursor, skeletal muscle - rabbit  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: A34154; S13047  
R.Fliegell, L.; Burns, K.; MacLennan, D.H.; Reithmeier, R.A.F.; Michalak, M.  
J. Biol. Chem. 264, 21522-21528, 1999  
A:Title: Molecular cloning of the high affinity calcium-binding protein (calreticulin) c  
A:Reference number: A34154; MUID:90094320; PMID:2600080  
A:Accession: A34154  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-418 <FL>  
A:Cross-references: GB:J05138; NID:G164858; PIDN:AAA1188.1; PID:G164859  
R:Reviews: S.; de Mattei, M.; Lantredi, M.; Villa, A.; Green, N.M.; MacLennan, D.H.; Meld  
Biochem. J. 271, 473-480, 1990  
A:Title: Calreticulin is a candidate for a calnexin-like function in Ca(2+)-storage  
A:Reference number: S13045; MUID:91054414; PMID:2241926  
A:Accession: S13047  
A:Molecule type: protein  
A:Residues: 19-32 <TR>  
C:Superfamily: calreticulin  
C:Keywords: skeletal muscle  
F:1-17/Domain: signal sequence #status predicted <SIG>  
F:415-418/Region: endoplasmic reticulum retention signal

Query Match 51.7%; Score 45; DB 1; Length 418;  
Best Local Similarity 83.3%; Pred. No. 23;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LPLLLGLGPA 13  
DB 5 VPLLLGLGLAA 16

RESULT 10  
D57073  
PTS system, beta-glucosides-specific IIABC component [imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C>Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C:Accession: D97073  
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.U.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: D97073  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-628 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AAK79375.1; PID:G15024346; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC1407  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase

Query Match 51.7%; Score 45; DB 2; Length 628;  
Best Local Similarity 50.0%; Pred. No. 33;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 MPELLGLGPA 16  
DB 264 WPELLGLGPA 279

RESULT 11  
A55429  
11-cis retinol dehydrogenase (EC 1.1.1.-) - bovine  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 10-Feb-1995 #sequence\_revision 10-Feb-1995 #text\_change 29-Sep-1999  
C:Accession: A55429  
R:Simon, A.; Hellman, U.; Wernstedt, C.; Eriksson, U.  
U. Biol. Chem. 270, 1107-1112, 1995  
A:Title: The retinal pigment epithelial-specific 11-cis retinol dehydrogenase belongs to  
A:Reference number: A55429; MUID:95138097; PMID:7836368  
A:Accession: A55429  
A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA  
A:Residues: 1-318 <STM>  
A:Cross-references: GB:X82262; NID:G663170; PIDN:CAA57715.1; PID:G663171  
C:Superfamily: retinol dehydrogenase; short-chain alcohol dehydrogenase homology  
C:Keywords: membrane protein; NAD; oxidoreductase  
F:29-206/Domain: short-chain alcohol dehydrogenase homology <SADH>

Query Match 50.6%; Score 44; DB 2; Length 318;  
Best Local Similarity 71.4%; Pred. No. 25;  
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 LPLLLGLGPA 15  
DB 3 LPLLLGLVLMALW 16

RESULT 12  
F75407  
probable transport protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: F75407  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.U.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Mah  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567286  
A:Accession: F75407  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-553 <MHI>  
A:Cross-references: GB:AE001980; GB:AE000513; NID:G6455086; PIDN:AAF10908.1; PID:G645908;  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR1336  
A:Map position: 1

Query Match 50.6%; Score 44; DB 2; Length 553;  
Best Local Similarity 58.3%; Pred. No. 42;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 LLLGLGPA 15  
DB 208 LLLGLGPA 219

RESULT 13  
T18291  
patched protein - zebra fish  
C:Species: Brachydanio rerio (zebra fish)  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000  
C:Accession: T18291  
R:Concordat, J.P.; Lewis, K.B.; Moore, J.W.; Goodrich, L.V.; Johnson, R.L.; Scott, M.P.;  
Development 122, 2835-2846, 1996  
A:Title: Spatial regulation of a zebrafish patched homologue reflects the roles of sonic  
A:Reference number: Z18860; MUID:96379744; PMID:8787757  
A:Accession: T18291  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1220 <CON>  
A:Cross-references: EMBL:X98883; PIDN:CAA67386.1  
C:Genetics:  
A:Gene: ptc1  
C:Superfamily: Drosophila membrane protein patched

Query Match 50.6%; Score 44; DB 2; Length 1220;  
Best Local Similarity 61.5%; Pred. No. 86;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MPELLGLGPA 13  
DB 1159 LPELLGLGPA 1171

## RESULT 14

C81919  
Probable membrane protein NMA0754 [imported] - Neisseria meningitidis (strain Z2491 sero  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: C81919  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
; Holroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A:Reference number: AB1775; MUID:2022556; PMID:10761919  
A:Accession: C81919  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-119 <PAR>  
A:Cross-references: GB:AL162754; GB:AL157959; NID:G7379424; PID: CAB84037.1; PID:G737947  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: NMA0754

## Query Match

49.4%; Score 43; DB 2; Length 119;

Best Local Similarity 34.5%; Pred. No. 15;

Matches 10; Conservative 3; Mismatches 2; Indels 14; Gaps 1;

OY 2 LPLILGLIG-----PACWA 16

Db 12 ISLLIGIGIFLPLPTTPFVLLSAACWA 40

## RESULT 15

F81182  
conserved hypothetical protein NMB0571 [imported] - Neisseria meningitidis (strain MCS8  
C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: F81182  
R:Teitelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
rt, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignan, V.; Piza, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MCS8.  
A:Reference number: AB1000; MUID:20175755; PMID:10710307  
A:Accession: F81182  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-119 <TER>  
A:Cross-references: GB:AE002413; GB:AE002098; NID:G7225797; PID:AAF40999.1; PID:G722579  
A:Experimental source: serogroup B, strain MCS8  
C:Genetics:  
A:Gene: NMB0571

## Query Match

49.4%; Score 43; DB 2; Length 119;

Best Local Similarity 34.5%; Pred. No. 15;

Matches 10; Conservative 3; Mismatches 2; Indels 14; Gaps 1;

OY 2 LPLILGLIG-----PACWA 16

Db 12 ISLLIGIGIFLPLPTTPFVLLSAACWA 40

Search completed: January 12, 2004, 08:16:49  
Job time : 3.92615 secs

**THIS PAGE BLANK (USPTO)**



01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ephrin type-B receptor 3 precursor (EC 2.7.1.112) (Tyrosine-protein  
 kinase receptor Mdk-5) (Developmental kinase 5) (SBK-4).  
 GN EPHB3 OR ETK2 OR MDK5 OR SEK4.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RX SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Embryo;  
 MDLINE=96074837; PubMed=7478528;  
 RA Ciosek T., Leach M.M., Ulrich A.;  
 RT "Cloning, characterization, and differential expression of MDK2 and  
 RT MDK5, two novel receptor tyrosine kinases of the eck/eph family.";  
 RT Oncogene 11:2085-2095(1995).  
 RL [2]  
 RN SEQUENCE OF 719-993 FROM N.A.  
 RP STRAIN=BALB/c; TISSUE=Brain;  
 RX MEDLINE=95034306; PubMed=7947319;  
 RA Becker N., Seitanidou T., Murphy P., Mattei M.-G., Topilko P.,  
 RA Niero A., Wilkinson D.G., Charney P., Giaraldi P.;  
 RT "Several receptor tyrosine kinase genes of the Eph family are  
 RT segmentally expressed in the developing hindbrain.";  
 RT Mech. Dev. 47:3-17(1994).  
 RL [3]  
 RN TISSUE SPECIFICITY.  
 RX MEDLINE=20171264; PubMed=10704386;  
 RA Mondini R., Wideman C., Kaprielian Z.;  
 RT "Complementary expression of transmembrane ephrins and their receptors  
 RT in the mouse spinal cord: a possible role in constraining the  
 RT orientation of longitudinally projecting axons.";  
 RT Development 127:1397-1410(2000).  
 RL -1- FUNCTION: RECEPTOR FOR MEMBERS OF THE EPHRIN-B FAMILY.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN CELLS OF THE RETINAL GANGLION  
 CC CELL LAYER DURING RETINAL AXON GUIDANCE TO THE OPTIC DISK.  
 CC -1- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.  
 CC -1- SIMILARITY: Contains 2 fibronectin type III domains.  
 CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. EPHRIN  
 CC RECEPTOR SUBFAMILY.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; 249086; CAAB8910.1; -  
 CC EMBL; X76012; CAAS3599.1; -  
 CC PIR; 148653; 148653.  
 CC PIR; 148761; 148761.  
 CC HSSP; P29323; 1B4F.  
 DR GMD; MGJ:104770; Ephb3.  
 DR GO; GO:0008046; P:axon guidance receptor activity; IDA.  
 DR GO; GO:0007411; P:axon guidance; IDA.  
 DR InterPro; IPR001090; Ephrin\_receptor.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR003962; FNIII subd.  
 DR InterPro; IPR000719; ProK Kinase.  
 DR InterPro; IPR001660; SAM.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR InterPro; IPR001426; YKase\_receptorV.  
 DR Pfam; PF01404; EPH\_Ibd; 1.  
 DR Pfam; PF00041; fn3; 2.  
 DR Pfam; PF00069; pkinae; 1.  
 DR Pfam; PF00536; SAM; 1.  
 DR PRINTS; PR00014; FNTYPEIII.

DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD001495; Ephrin\_receptor; 1.  
 DR ProDom; PD000001; Prot\_Kinase; 1.  
 DR SMART; SM00615; EPH\_Ibd; 1.  
 DR SMART; SM00060; FN3; 2.  
 DR SMART; SM00454; SAM; 1.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS00790; RECEPTOR\_TYR\_KIN\_V\_1; 1.  
 DR PROSITE; PS00791; RECEPTOR\_TYR\_KIN\_V\_2; 1.  
 DR PROSITE; PS00105; SAM\_DOMAIN; 1.  
 DR Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KW Receptor; Transmembrane; Glycoprotein; Signal; Repeat.  
 FT SIGNAL 1 29  
 FT CHAIN 30 993  
 FT DOMAIN 30 554  
 FT TRANSMEM 555 575  
 FT DOMAIN 576 993  
 FT DOMAIN 191 328  
 FT DOMAIN 329 443  
 FT DOMAIN 444 539  
 FT DOMAIN 628 891  
 FT DOMAIN 920 984  
 FT SITE 991 993  
 FT NP\_BIND 634 642  
 FT BINDING 660 660  
 FT ACT\_SITE 753 753  
 FT MOD\_RES 603 603  
 FT MOD\_RES 609 609  
 FT MOD\_RES 787 787  
 FT MOD\_RES 937 937  
 FT CARBOHYD 343 343  
 FT CARBOHYD 440 440  
 FT CONFLICT 719 719  
 SQ SEQUENCE 993 AA; 109585 MW; 0B66A4D391266C79 CRC64;  
 Query Match 65.5%; Score 57; DB 1; Length 993;  
 Best Local Similarity 68.8%; Pred. No. 0.42;  
 Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 ID 1 MPTLLGLLPACWA 16  
 DB 14 LAPLLPLLPACWA 29  
 RESULT 3  
 PC12 MOUSE STANDARD; PRT; 1180 AA.  
 AC 055134;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DE Protocadherin 12 precursor (Vascular cadherin-2)  
 DE cadherin-2 (VE-cadherin-2) (VE-cad-2).  
 GN PCDH12.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RX SEQUENCE FROM N.A.  
 RC TISSUE=Brain capillary;  
 MDLINE=98316322; PubMed=9651350;  
 RA Telo P., Breviario F., Huber P., Panzeri C., Dejana E.;  
 RT "Identification of a novel cadherin (vascular endothelial cadherin-2)  
 RT located at intercellular junctions in endothelial cells.";  
 RL J. Biol. Chem. 273:17565-17572(1998).  
 CC -1- FUNCTION: Cellular adhesion molecule that may play an important  
 CC role in cell-cell interactions at interendothelial junction.  
 CC Promotes homotypic calcium dependent aggregation and adhesion and  
 CC clusters at intercellular junctions. Unable to bind to catenins,



weakly associates with the cytoskeleton.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: Expressed in endothelial cells.  
 CC -1- SIMILARITY: Contains 6 cadherin domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Y08715; CA669965.1; -  
 DR PIR: T31066; T31066.  
 DR MGD: MGI:1855700; Pcdh12.  
 DR GO: GO:0005911; C:intracellular junction; IDA.  
 DR GO: GO:0008014; F:calcium-dependent cell adhesion molecule ac. .; IDA.  
 DR GO: GO:0007156; P:homophilic cell adhesion; IDA.  
 DR InterPro: IPR002126; Cadherin.  
 DR Pfam: PF00028; cadherin.5.  
 DR PRINTS: PR00205; CADHERIN.  
 DR SMART: SM00112; CA; 6.  
 DR PROSITE: PS00232; CADHERIN 1; 5.  
 DR PROSITE: PS50268; CADHERIN 2; 6.  
 DR Transmembrane: Glycoprotein; Repeat; Cell adhesion; Calcium;  
 KW Cytoskeleton; Signal.  
 KM SIGNAL 1 17  
 FT CHAIN 18 1180  
 FT DOMAIN 1 716  
 FT TRANSMEM 717 737  
 FT DOMAIN 738 1180  
 FT DOMAIN 28 135  
 FT DOMAIN 136 244  
 FT DOMAIN 245 352  
 FT DOMAIN 355 460  
 FT DOMAIN 461 565  
 FT DOMAIN 566 711  
 FT CARBOHYD 265 265  
 FT CARBOHYD 415 415  
 FT CARBOHYD 582 582  
 FT CARBOHYD 659 659  
 FT CARBOHYD 662 662  
 FT SEQUENCE 1180 AA; 128673 MW; 97927ADP1FC38087 CRC64;  
 Query Match 56.3%; Score 49; DB 1; Length 1180;  
 Best Local Similarity 69.2%; Pred. No. 7.9;  
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 LPLLLGLGPPA 13  
 Db 4 LPLLLGLGPPG 16  
 RESULT 4  
 ID CRIC\_MOUSE STANDARD; PRT; 416 AA.  
 AC P14211;  
 DT 01-JAN-1990 (Rel. 13; Created)  
 DT 01-JAN-1990 (Rel. 13; Last sequence update)  
 DT 15-DEC-1998 (Rel. 37; Last annotation update)  
 DE Calreticulin precursor (CRP55) (Calregulin) (HACBP) (ERP60).  
 GN CALR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 18-48 AND 129-161.  
 RP STRAIN=BALB/c; TISSUE=Liver;  
 RX MEDLINE=90059955; PubMed=2583110;  
 RA Smith M.J., Koch G.L.E.;  
 RT "Multiple zones in the sequence of calreticulin (CRP55, calregulin,

HACBP), a major calcium binding ER/SR protein.";  
 RT EMBL J. 8:3581-3586(1989).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93013037; PubMed=1398135;  
 RA Mazarella R.A., Gold P., Cunningham M., Green M.;  
 RT "Determination of the sequence of an expressible cDNA clone encoding  
 RT ERp60/calregulin by the use of a novel nested set method.";  
 RL Gene 120:217-225(1992).  
 RN [3]  
 RP SEQUENCE OF 18-38.  
 RC TISSUE=Fibroblast;  
 RX MEDLINE=95009907; PubMed=7523108;  
 RA Merrick B.A., Patterson R.M., Wichter L.J., He C., Selkirk J.K.;  
 RT "Separation and sequencing of familial and novel murine proteins  
 RT using preparative two-dimensional gel electrophoresis.";  
 RL Electrophoresis 15:735-745(1994).  
 CC -1- FUNCTION: THIS PROTEIN BINDS CALCIUM. THERE ARE BOTH HIGH AND  
 CC LOW AFFINITY CALCIUM-BINDING SITES.  
 CC -1- SUBUNIT: Monomer (by similarity).  
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.  
 CC -1- SIMILARITY: BELONGS TO THE CALRETICULIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X14926; CA333053.1; -  
 DR EMBL: M92988; AAA37569.1; -  
 DR PIR: S06763; S06763.  
 DR SWISS-2DPAGE: P14211; MOUSE.  
 DR MGD: MGI:88252; Calr.  
 DR InterPro: IPR001580; Calreticulin.  
 DR InterPro: IPR000886; ER target.  
 DR Pfam: PF00262; calreticulin.1.  
 DR PRINTS: PR00626; CALRETICULIN.  
 DR PRODOM: PD001866; Calreticulin; 1.  
 DR PROSITE: PS0014; ER TARGET; 1.  
 DR PROSITE: PS00803; CALRETICULIN 1; 1.  
 DR PROSITE: PS00804; CALRETICULIN 2; 1.  
 DR PROSITE: PS00805; CALRETICULIN\_REPEAT; 3.  
 DR Endoplasmic reticulum; Calcium-binding; Repeat; Signal.  
 KM SIGNAL 1 17  
 FT CHAIN 18 416  
 FT DOMAIN 18 197  
 FT DOMAIN 198 308  
 FT DOMAIN 309 416  
 FT DOMAIN 191 255  
 FT REPEAT 191 202  
 FT REPEAT 210 221  
 FT REPEAT 227 238  
 FT REPEAT 244 255  
 FT REPEAT 259 297  
 FT REPEAT 259 269  
 FT REPEAT 273 283  
 FT REPEAT 287 297  
 FT DOMAIN 351 407  
 FT DISULFID 137 163  
 FT SITE 413 416  
 FT SEQUENCE 416 AA; 47994 MW; 24C03B0913408D8 CRC64;  
 Query Match 51.7%; Score 45; DB 1; Length 416;  
 Best Local Similarity 83.3%; Pred. No. 13;  
 Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Oy 2 LPLLLGLGPPA 13  
 Db 5 VPLLLGLGLAA 16

RESULT 5  
 CRTC\_RAT STANDARD; PRT; 416 AA.  
 AC P18418; P10452; (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 01-NOV-1990 (Rel. 16, Last annotation update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Calreticulin precursor (CRP55) (Calregulin) (HACBP) (ERP60) (CALBP)  
 DE (Calcium-binding protein 3) (CABP3).  
 GN CALR.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OK NCBI\_TaxID=10116;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Brain cortex;  
 RX MEDLINE=90370496; PubMed=2395661;  
 RA Holmes C., Banville D., Strikant C.B., Carrier F., Bell A.,  
 RA Murphy K.K., Patel Y.C.;  
 RA "Structural homology between the rat calreticulin gene product and  
 RT the Oncoferrin volvulus antigen Ral-1.";  
 RL Nucleic Acids Res. 18:4933-4933(1990).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RX MEDLINE=93302172; PubMed=8453984;  
 RA Nakamura M., Moriya M., Baba T., Michikawa Y., Yamanobe T., Arai K.,  
 RA Okinaga S., Kobayashi T.;  
 RA "An endoplasmic reticulum protein, calreticulin, is transported into  
 RT the endosome of rat sperm.";  
 RL Exp. Cell Res. 205:101-110(1993).  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
 RX MEDLINE=95181573; PubMed=7876339;  
 RA Soennichsen B., Fuellekrug J., van Nguyen P., Diekmann W.,  
 RA Robinson D.G., Mieskes G.;  
 RA "Retention and retrieval: both mechanisms cooperate to maintain  
 RT calreticulin in the endoplasmic reticulum.";  
 RL J. Cell Sci. 107:2705-2717(1994).  
 RN (4)  
 RP SEQUENCE OF 270-358 FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RX Lone Y.C., Bailly A., Lacroffe N.;  
 RL Submitted (DEC-1988) to the EMBL/GenBank/DBJ databases.  
 RN (5)  
 RP SEQUENCE OF 18-29.  
 RX MEDLINE=91054414; PubMed=2241926;  
 RA Treves S., de Mattei M., Landfredi M., Villa A., Green N.M.,  
 RA MacLennan D.H., Meldolesi J., Pozzan T.;  
 RA "Calreticulin is a candidate for a calsequestrin-like function in  
 RT Ca2(+)-storage compartments (calciosomes) of liver and brain.";  
 RL Biochem. J. 271:473-480(1990).  
 RN (6)  
 RP SEQUENCE OF 18-32.  
 RC STRAIN=Sprague-Dawley; TISSUE=Testis;  
 RX MEDLINE=92360010; PubMed=1497655;  
 RA Nakamura M., Michikawa Y., Baba T., Okinaga S., Arai K.;  
 RA "Calreticulin is present in the endosome of spermatids of rat  
 RT testis.";  
 RL Biochem. Biophys. Res. Commun. 186:668-673(1992).  
 RN (7)  
 RP SEQUENCE OF 18-32.  
 RC STRAIN=LEC; TISSUE=Liver;  
 RX MEDLINE=94072621; PubMed=8251535;  
 RA Yokoi T., Nagayama S., Kajiwara R., Kawaguchi Y., Horuchi R.,  
 RA Kametaki T.;  
 RA "Identification of protein disulfide isomerase and calreticulin as  
 RT autoimmune antigens in LEC strain of rats.";  
 RL Biochim. Biophys. Acta 1158:339-344(1993).  
 CC -1- FUNCTION: THIS PROTEIN BINDS CALCIUM. THERE ARE BOTH HIGH AND

CC LOW AFINITY CALCIUM-BINDING SITES.  
 CC -1- SUBUNIT: Monomer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.  
 CC -1- SIMILARITY: BELONGS TO THE CALRETICULIN FAMILY.  
 CC -1- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE D-BETA-HYDROXYBUTYRATE DEHYDROGENASE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D78308; BAA11345.1; -;  
 DR EMBL; X53363; CAA37446.1; -;  
 DR EMBL; X13702; CAA31987.1; ALT\_SEQ.  
 DR EMBL; X79327; CAA55890.1; -;  
 DR PIR; JH0819; JH0819.  
 DR PDB; 1H9N; 26-FEB-02.  
 DR PDB; 1K91; 12-OCT-02.  
 DR PDB; 1K9C; 12-OCT-02.  
 DR InterPro; IPR001580; Calreticulin.  
 DR InterPro; IPR000886; ER\_target.  
 DR Pfam; PF00262; calreticulin; 1.  
 DR PRINTS; PR00626; CALRETICULIN.  
 DR PRODOM; PD001866; Calreticulin; 1.  
 DR PROSITE; PS00014; ER\_TARGET; 1.  
 DR PROSITE; PS00803; CALRETICULIN\_1; 1.  
 DR PROSITE; PS00804; CALRETICULIN\_2; 1.  
 DR PROSITE; PS00805; CALRETICULIN\_REPEAT; 3.  
 DR KX Endoplasmic reticulum; Calcium-binding; Repeat; Signal; 3D-structure.  
 FT SIGNAL 1 17  
 FT CHAIN 18 416 CALRETICULIN.  
 FT DOMAIN 18 197 N-DOMAIN.  
 FT DOMAIN 198 308 P-DOMAIN.  
 FT DOMAIN 309 416 C-DOMAIN.  
 FT DOMAIN 191 255 4 X APPROXIMATE REPEATS.  
 FT REPEAT 191 202 1-1.  
 FT REPEAT 210 221 1-2.  
 FT REPEAT 227 238 1-3.  
 FT REPEAT 244 255 1-4.  
 FT DOMAIN 259 297 3 X APPROXIMATE REPEATS.  
 FT REPEAT 259 269 2-1.  
 FT REPEAT 273 283 2-2.  
 FT REPEAT 351 407 2-3.  
 FT DOMAIN 137 163 ASP/GLU/LYS-RICH.  
 FT DISULFID 137 163 BY SIMILARITY.  
 FT SITE 413 416 PREVENT SECRETION FROM ER.  
 SQ SEQUENCE 416 AA; 47995 MW; 2B6713CEB31A2970 CRC64;  
 Query Match 51.7%; Score 45; DB 1; Length 416;  
 Best Local Similarity 83.3%; Pred. No. 13;  
 Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 LPILGLIGPPA 13  
 DB 5 VPLLGLGLAA 16  
 RESULT 6  
 CRTL\_BOVIN STANDARD; PRT; 417 AA.  
 AC P52193; Q8SQ53; (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Calreticulin, brain isoform 1 precursor (CRP55) (Calregulin) (HACBP).  
 GN CALR OR CRT.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Hosain M.A.; Takawa K.; Minakata H.; Nakajima T.;  
RT "Bovine brain calreticulin";  
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 18-417.  
RC TISSUE=Brain;  
RA MEDLINE=94183174; PubMed=8135753;  
RA Matsuno K.; Seta K.; Yamakawa Y.; Okuyama T.; Shinoda T.; Isebe T.;  
RT "Covalent structure of bovine brain calreticulin";  
RL Biochem J 298:435-442(1994).  
CC -1- FUNCTION: THIS PROTEIN BINDS CALCIUM. THERE ARE BOTH HIGH AND  
CC -1- LOW AFFINITY CALCIUM-BINDING SITES.  
CC -1- SUBUNIT: Monomer.  
CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.  
CC -1- SIMILARITY: BELONGS TO THE CALRETICULIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AB067687; BAB86913.1; -  
DR InterPro; IPR001580; Calreticulin.  
DR InterPro; IPR000886; ER target.  
DR Pfam; PF00262; calreticulin; 1.  
DR PRINTS; PR00626; CALRETICULIN.  
DR PRODOM; PD001866; Calreticulin; 1.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
DR PROSITE; PS00803; CALRETICULIN\_1; 1.  
DR PROSITE; PS00804; CALRETICULIN\_2; 1.  
DR PROSITE; PS00805; CALRETICULIN\_REPEAT; 3.  
DR PROSITE; PS00805; CALRETICULIN\_REPEAT; 3.  
KM Endoplasmic reticulum; Calcium-binding; Repeat; Glycoprotein; Signal.  
FT SIGNAL 1 17  
FT CHAIN 18 417  
FT DOMAIN 18 197  
FT DOMAIN 198 308  
FT DOMAIN 309 417  
FT DOMAIN 191 255  
FT DOMAIN 191 202  
FT REPEAT 210 221  
FT REPEAT 227 238  
FT REPEAT 244 255  
FT DOMAIN 259 297  
FT REPEAT 259 269  
FT REPEAT 273 283  
FT REPEAT 287 297  
FT DOMAIN 351 407  
FT DISULFID 137 163  
FT CARBOHD 179 179  
FT SITE 414 417  
SQ SEQUENCE 417 AA; 48038 MW; 7BF812C7B5417BE9 CRC64;  
Query Match 51.7%; Score 45; DB 1; Length 417;  
Best Local Similarity 83.3%; Pred. No. 13;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Calreticulin precursor (CRP55) (Calregulin) (HACBP) (ERp60).  
GN CALR.  
OS Cricetus griseus (Chinese hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OC Cricetus.  
OX NCBI\_TaxID=10029;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Chung J.-Y.; Lee G.M.;  
RT Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.  
RL -1- FUNCTION: THIS PROTEIN BINDS CALCIUM. THERE ARE BOTH HIGH AND  
CC -1- LOW AFFINITY CALCIUM-BINDING SITES.  
CC -1- SUBUNIT: Monomer.  
CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.  
CC -1- SIMILARITY: BELONGS TO THE CALRETICULIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AY100688; AAM48568.1; -  
DR InterPro; IPR001580; Calreticulin.  
DR InterPro; IPR000886; ER target.  
DR Pfam; PF00262; calreticulin; 1.  
DR PRINTS; PR00626; CALRETICULIN.  
DR PRODOM; PD001866; Calreticulin; 1.  
DR PROSITE; PS00803; CALRETICULIN\_1; 1.  
DR PROSITE; PS00804; CALRETICULIN\_2; 1.  
DR PROSITE; PS00805; CALRETICULIN\_REPEAT; 3.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
KM Endoplasmic reticulum; Calcium-binding; Repeat; Glycoprotein; Signal.  
FT SIGNAL 1 17  
FT CHAIN 18 417  
FT DOMAIN 18 197  
FT DOMAIN 198 308  
FT DOMAIN 309 417  
FT DOMAIN 191 255  
FT DOMAIN 191 202  
FT REPEAT 210 221  
FT REPEAT 227 238  
FT REPEAT 244 255  
FT DOMAIN 259 297  
FT REPEAT 259 269  
FT REPEAT 273 283  
FT REPEAT 287 297  
FT DOMAIN 351 407  
FT DISULFID 137 163  
FT SITE 414 417  
SQ SEQUENCE 417 AA; 48242 MW; D617DA37D14F2D45 CRC64;  
Query Match 51.7%; Score 45; DB 1; Length 417;  
Best Local Similarity 83.3%; Pred. No. 13;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

RESULT 8  
CRTC\_RABIT  
ID CRTC\_RABIT STANDARD; PRT; 418 AA.  
AC P15253;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)

01-OCT-1996 (Rel. 34, Last annotation update)  
 Calreticulin precursor (CRP55) (Calreticulin) (HACBP) (ERp60).  
 GN CALR.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OK NCBI\_TaxID=9986;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=slow-twitch skeletal muscle;  
 RX MEDLINE=90094320; PubMed=2600080;  
 RA Fliegel L., Burns K., MacLennan D.H., Reithmeier R.A.F., Michalak M.;  
 RT "Molecular cloning of the high affinity calcium-binding protein  
 (calreticulin) of skeletal muscle sarcoplasmic reticulum.";  
 RL J. Biol. Chem. 264:21522-21528 (1989).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=fast-twitch skeletal muscle;  
 RX MEDLINE=91282795; PubMed=2059224;  
 RA Fliegel L., Michalak M.;  
 RT "Fast-twitch and slow-twitch skeletal muscles express the same  
 isoform of calreticulin.";  
 RL Biochem. Biophys. Res. Commun. 177:979-984 (1991).  
 RN (3)  
 RP SEQUENCE OF 18-36.  
 RX MEDLINE=91054414; PubMed=2241926;  
 RA Treves S., de Mattei M., Lanfredi M., Villa A., Green N.M.,  
 RA MacLennan D.H., Meldolesi J., Pozzan T.;  
 RT "Calreticulin is a candidate for a calsequestrin-like function in  
 Ca2(+)-storage compartments (calciosomes) of liver and brain.";  
 RL Biochem. J. 271:473-480 (1990).  
 RN (4)  
 RP SEQUENCE OF 18-46.  
 RX MEDLINE=91201375; PubMed=2016321;  
 RA Milner R.E., Baksh S., Shemanko C., Carpenter M.R., Smillie L.,  
 RA Vance J.E., Opas M., Michalak M.;  
 RT "Calreticulin, and not calsequestrin, is the major calcium binding  
 protein of smooth muscle sarcoplasmic reticulum and liver endoplasmic  
 reticulum.";  
 RL J. Biol. Chem. 266:7155-7165 (1991).  
 RN (5)  
 RP PARTIAL SEQUENCE.  
 RC TISSUE=Lung.  
 RX MEDLINE=92002038; PubMed=1911780;  
 RA Guan S., Fallick A.M., Williams D.E., Cashman J.R.;  
 RT "Evidence for complex formation between rabbit lung flavin-containing  
 monooxygenase and calreticulin.";  
 RL Biochemistry 30:9892-9900 (1991).  
 CC -1- FUNCTION: THIS PROTEIN BINDS CALCIUM. THERE ARE BOTH HIGH AND  
 LOW AFFINITY CALCIUM-BINDING SITES.  
 CC -1- SUBUNIT: Monomer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.  
 CC -1- SIMILARITY: BELONGS TO THE CALRETICULIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: J05138; AAA31188.1; -  
 DR PIR: A34154; A34154.  
 DR InterPro: IPR001580; Calreticulin.  
 DR InterPro: IPR000866; ER target.  
 DR Pfam: PF00262; calreticulin.1.  
 DR PRINTS: PR00626; CALRETICULIN.  
 DR ProDom: PD001866; Calreticulin; 1.  
 DR PROSITE: PS00014; ER\_TARGET; 1.  
 DR PROSITE: PS00803; CALRETICULIN\_1; 1.  
 DR PROSITE: PS00804; CALRETICULIN\_2; 1.  
 DR PROSITE: PS00805; CALRETICULIN\_REPEAT; 3.

KM Endoplasmic reticulum; Calcium-binding; Repeat; Signal.  
 FT SIGNAL 1 17  
 FT CHAIN 18 418  
 FT DOMAIN 18 197  
 FT DOMAIN 198 308  
 FT DOMAIN 309 418  
 FT DOMAIN 191 255  
 FT REPEAT 191 202  
 FT REPEAT 210 221  
 FT REPEAT 227 238  
 FT REPEAT 244 255  
 FT DOMAIN 259 297  
 FT REPEAT 259 269  
 FT REPEAT 273 283  
 FT REPEAT 287 297  
 FT DOMAIN 351 408  
 FT DISULFID 137 163  
 FT SITE 415 418  
 FT VARIANT 35 35  
 FT CONFLICT 90 90  
 SQ SEQUENCE 418 AA; 48275 MW; B60828689DC763A6 CRC64;  
 Query Match 51.7%; Score 45; DB 1; Length 418;  
 Best Local Similarity 83.3%; Pred. No. 13;  
 Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 LPLILGLIGPPA 13  
 Db 5 VPLILGLIGLAA 16  
 RESULT 9  
 ENV\_SRV2  
 ID ENV\_SRV2 STANDARD; PRT; 574 AA.  
 AC P51515;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE ENV polypeptide (coat polypeptide) [contains: Coat protein GP70; Coat  
 protein GP20].  
 GN ENV.  
 OS Simian retrovirus SRV-2.  
 OC Viruses; Retroid viruses; Retroviridae; Betaretrovirus.  
 OK NCBI\_TaxID=39068;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8715131; PubMed=2435057;  
 RA Thayer R.M., Power M.D., Bryant M.L., Gardner M.B., Barr P.J.;  
 RA Luciw P.A.;  
 RT "Sequence relationships of type D retroviruses which cause simian  
 acquired immunodeficiency syndrome.";  
 RL Virology 157:317-329 (1987).  
 CC -1- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: M16605; AAA47563.1; ALT\_INIT.  
 DR HSSP: P03385; IMOF.  
 DR InterPro: IPR002050; Env polypeptide.  
 DR Pfam: PF00429; ENV polypeptide; 1.  
 KM Coat protein; Glycoprotein; Polypeptide; Transmembrane.  
 FT CHAIN 1 382  
 FT CHAIN 383 574  
 FT TRANSMEM 386 406  
 FT TRANSMEM 515 535  
 FT CARBOHYD 117 117  
 FT CARBOHYD 233 233  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL)  
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL)  
FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL)  
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL)  
FT CARBOHYD 312 312 N-LINKED (GLCNAC. . .) (POTENTIAL)  
FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL)  
FT CARBOHYD 327 327 N-LINKED (GLCNAC. . .) (POTENTIAL)  
FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL)  
FT CARBOHYD 475 475 N-LINKED (GLCNAC. . .) (POTENTIAL)  
SQ SEQUENCE 574 AA; 62952 MW; 662B1E1A437B055 CRC64;  
Query Match 51.7%; Score 45; DB 1; Length 574;  
Best Local Similarity 64.3%; Pred. No. 17;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 MLPLLLGLLGPAC 14  
Db 510 LPLVLLPLGLPLFC 523  
RESULT 10  
RDH1 BOVIN STANDARD; PRT; 318 AA.  
AC Q27979;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 11-cis retinol dehydrogenase (EC 1.1.1.105) (11-cis RDH) (P32).  
GN RDH5 OR RDH1.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=EYE;  
RX MEDLINE=95138997; PubMed=7836368;  
RA Simon A., Hellman U., Wernstedt C., Eriksson U.;  
RT "The retinal pigment epithelial-specific 11-cis retinol dehydrogenase  
belongs to the family of short chain alcohol dehydrogenases.";  
J. Biol. Chem. 270:1107-1112(1995).  
RL -1- FUNCTION: STEREOSPECIFIC 11-CIS RETINOL DEHYDROGENASE, WHICH  
CATALYZES THE FINAL STEP IN THE BIOSYNTHESIS OF 11-CIS  
RETINALDEHYDE, THE UNIVERSAL CHROMOPHORE OF VISUAL PIGMENTS.  
CC ACTIVE IN THE PRESENCE OF NAD AS COFACTOR BUT NOT IN THE PRESENCE  
CC OF NADP.  
CC -1- CATALYTIC ACTIVITY: Retinol + NAD(+) = retinal + NADH.  
CC -1- PATHWAY: Retinoic acid biosynthesis; first (rate-limiting) step.  
CC -1- SUBCELLULAR LOCATION: Membrane-associated.  
CC -1- SIMILARITY: Belongs to the short-chain dehydrogenases/reductases  
CC (SDR) family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL, X82262; CAA57715.1; -;  
CC PIR, A55429; A55429.  
CC HSP, P14061; 1PDU.  
CC InterPro: IPR002198; ADH short.  
CC Pfam: PF00106; adh short; 1.  
CC PRINTS: PR00080; SDRFAMILY.  
CC PROSITE, PS00061; ADH\_SHORT; FALSE\_NEG.  
CC Oxidoreductase; NAD; Membrane; Vision.  
KW ACT SITE 32  
FT NADP (BY SIMILARITY).  
FT AC BIND 175 56  
FT BY SIMILARITY.  
SQ SEQUENCE 318 AA; 35036 MW; 78B4A1E435F351 CRC64;

Query Match 50.6%; Score 44; DB 1; Length 318;  
Best Local Similarity 71.4%; Pred. No. 15;  
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
Qy 2 LPLLLGLLGPACW 15  
Db 3 LPLLLGLLGPACW 16  
RESULT 11  
PC12 HUMAN STANDARD; PRT; 1184 AA.  
AC Q9NGG4; Q9GKB8; Q9HY6; Q9H8E0;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Procadherin 12 precursor (Vascular cadherin-2) (Vascular endothelial  
cadherin-2) (VE-cadherin-2) (VE-cad-2).  
GN PCDH12.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20202599; PubMed=10716726;  
RA Wu Q., Maniatis T.;  
RT "Large exons encoding multiple ectodomains are a characteristic  
feature of protocadherin genes.";  
Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).  
RL [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Retinal Kidney;  
RX MEDLINE=20515266; PubMed=11063261;  
RA Ludwig D., Lorenz T., Dejana E., Bohnen P., Hicklin D.J., Witte L.,  
RA Pytkowski B.;  
RT "cDNA cloning, chromosomal mapping, and expression analysis of human  
VE-cadherin-2.";  
Mamm. Genome 11:1030-1033(2000).  
RL [3]  
RP SEQUENCE FROM N.A.  
RA Seki N., Hattori A., Hayashi A., Kozuma S., Muramatsu M., Saito T.;  
RT "Human vascular cadherin-2.";  
Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo, and Placenta;  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
RA Wagaetsuma M., Hosokiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Matanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuko Y.,  
RA Niimura K., Iwayanagi T.;  
RT "NEO human cDNA sequencing project.";  
Submitted (May-2001) to the EMBL/GenBank/DBJ databases.  
RL -1- FUNCTION: Cellular adhesion molecule that may play an important  
CC role in cell-cell interactions at interendothelial junctions.  
CC Promotes homotypic calcium dependent aggregation and adhesion and  
CC clusters at intercellular junctions. Unable to bind to catenins,  
CC weakly associates with the cytoskeleton (By similarity).  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- TISSUE SPECIFICITY: Expressed in highly vascularized tissues  
CC including the heart and placenta, but most tissues contain a low  
CC level of expression. Prominent expression in the spleen.  
CC -1- SIMILARITY: Contains 6 cadherin domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>



Query Match 50.6%; Score 44; DB 1; Length 1220;  
 Best Local Similarity 61.5%; Pred. No. 46;  
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MLPLLLGLGPA 13  
 Db 1159 LLPVLISLMGPPA 1171

## RESULT 13

TAM\_PSEAE STANDARD; PRT; 275 AA.  
 AC Q910S1  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Trans-aconitate 2-methyltransferase (EC 2.1.1.144).  
 GN TAM OR PA2564.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 CC Pseudomonadaceae; Pseudomonas.  
 CC NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Eryan A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Gader R.L., Goltzy L., Tolentino E., Westbrock-Medman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lam R.M.,  
 RA Smith K.A., Spencer D.H., Wong K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 CC -1- FUNCTION: Catalyzes the S-adenosylmethionine monomethyl  
 CC esterification of trans-aconitate (By similarity).  
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + trans-aconitate = S-  
 CC adenosyl-L-homocysteine + (E)-3-(methoxycarbonyl)pent-2-enedioate.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE METHYLTRANSFERASE SUPERFAMILY. TAM  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AE004684; AAG05952.1; -  
 CC PIR; H83325; H83325.  
 CC HAMAP; MF\_00560; -; 1.  
 CC InterPro; IPR000051; SAM bind.  
 CC Transferrase; Methyltransferase; Complete proteome.  
 CC SEQUENCE 275 AA; 31674 MW; A902E2B71FF2E06B CRC64;  
 Qy Query Match 49.4%; Score 43; DB 1; Length 275;  
 Best Local Similarity 56.2%; Pred. No. 18;  
 Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MLPLLLGLGPAACWA 16  
 Db 124 LLPRLAGLTFGGCIA 139

RESULT 14  
 MIAA\_BUCBP STANDARD; PRT; 323 AA.  
 AC P59507;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE tRNA delta(2)-isopentenylpyrophosphate transferase (EC 2.5.1.8) (IPP  
 DE transferase) (isopentenyl-diphosphate:tRNA isopentenyltransferase)  
 DE (IPTase) (IPTT).  
 GN MIAA OR BPP514.  
 OS Buchnera aphidicola (subsp. Baizongia pistaciae).  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Buchnera.  
 CC NCBI\_TaxID=135842;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426901; PubMed=12522265;  
 RA Van Ham R.C.H.J., Kamebeek J., Palacios C., Rausell C., Abascal F.,  
 RA Bastolla U., Fernandez J.M., Jimenez L., Postigo M., Silva F.J.,  
 RA Tamames J., Viguera E., Latorre A., Valencia A., Moran F., Moya A.;  
 RT "Reductive genome evolution in Buchnera aphidicola.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:581-586(2003).  
 CC -1- FUNCTION: Catalyzes the first step in the biosynthesis of 2-  
 CC methylthio-N6-(delta(2)-isopentenyl)-adenosine (MS[2]16[A])  
 CC adjacent to the anticodon of several tRNA species (By similarity).  
 CC -1- CATALYTIC ACTIVITY: isopentenyl diphosphate + tRNA = diphosphate +  
 CC tRNA containing 6-isopentenyladenosine.  
 CC -1- SIMILARITY: BELONGS TO THE IPP TRANSFERASE FAMILY.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AE014017; AA027217.1; -  
 DR HAMAP; MF\_00185; -; 1.  
 DR Transferrase; Nucleotidyltransferase; tRNA processing; ATP-binding;  
 KW Complete proteome.  
 KM NP BIND 21  
 FT SEQUENCE 323 AA; 37553 MW; D6CFB62D90E1C9D CRC64;

Qy Query Match 49.4%; Score 43; DB 1; Length 323;  
 Best Local Similarity 58.3%; Pred. No. 21;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 3 PLLLGLGPAAC 14  
 Db 14 PLVIFLMGPTAC 25

## RESULT 15

COA2\_POVMK STANDARD; PRT; 341 AA.  
 ID COA2\_POVMK  
 AC P24596;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Coat protein VP2 [Contains: Coat protein VP3].  
 OS Mouse polyomavirus (strain Kilham).  
 CC Viruses; dsDNA viruses, no RNA stage; Polyomaviridae; Polyomavirus.  
 CC NCBI\_TaxID=10638;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91196237; PubMed=1849675;  
 RA Mayer M., Doerries K.;  
 RT "Nucleotide sequence and genome organization of the murine  
 RT polyomavirus, Kilham strain.";  
 RL Virology 181:469-480(1991).  
 RN [2]  
 RP REVISIONS TO 58 AND 121.  
 RA Mayer M., Doerries K.;  
 RL Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative initiation;  
 CC Comment=2 isoforms, VP2 (shown here) and VP3, are produced by

```

CC      alternative initiation;
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; M55904; AAA46554.1; -.
CC      EMBL; M55904; AAA46555.1; -.
CC      InterPro: IPR001070; Polyoma.coat2.
CC      Pfam; PR00761; Polyoma.coat2.1.
CC      Late protein; Coat protein; Alternative initiation.
CC      CHAIN 1 341
CC      FT CHAIN 120 341 COAT PROTEIN VP2, ISOFORM VP2.
CC      FT INIT MET 120 120 COAT PROTEIN VP2, ISOFORM VP3.
CC      FT CONFLICT 58 58 R -> A (IN REF. 1).
CC      FT CONFLICT 121 121 A -> V (IN REF. 1).
CC      SQ SEQUENCE 341 AA: 37396 MW: 9F8E3D25514BBF75 CRC64;

Query Match 49.4%; Score 43; DB 1; Length 341;
Best Local Similarity 90.0%; Pred. No. 22;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 MRLPLGLLG 10
        |||||
        283 MRLPLGLYG 292

```

Search completed: January 12, 2004, 08:13:01  
 Job time : 2.57465 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:10 ; Search time 2.20359 Seconds

(without alignments)  
1873.686 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_1\_16

Sequence: 1 MIPLLGLGPACWA 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

SPREMBL\_23.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriophage.\*  
17: sp\_archaeal.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	100.0	329	4	Q8IV25
2	73	83.9	326	11	Q8R066
3	57	65.5	993	11	Q91Y89
4	52	59.8	115	16	Q51537
5	49	56.3	155	16	Q8PCRO
6	49	56.3	998	11	Q8C883
7	49	56.3	1061	5	Q9V9L4
8	48	55.2	379	11	Q9DCW1
9	47	54.0	569	8	Q94RB8
10	46.5	53.4	719	16	Q8P3B9
11	46	52.9	228	16	Q9AK63
12	46	52.9	496	5	Q967F2
13	46	52.9	517	11	Q8CH08
14	46	52.9	517	11	Q8B148
15	46	52.9	558	3	Q9UTM5
16	46	52.9	580	15	Q9WP29

17	45	51.7	309	16	Q8UY5	Q8UY5 agrobacteri
18	45	51.7	315	16	Q9RU53	Q9RU53 deinococcus
19	45	51.7	347	4	Q9Y6U7	Q9Y6U7 homo sapien
20	45	51.7	395	16	Q8PF13	Q8PF13 xanthomonas
21	45	51.7	488	5	Q9NI45	Q9NI45 trypanosoma
22	45	51.7	602	5	Q9YUJ3	Q9YUJ3 drosophila
23	45	51.7	628	16	Q9YUJ9	Q9YUJ9 clostridium
24	44.5	51.1	232	12	Q89737	Q89737 polyomaviru
25	44.5	51.1	253	16	Q8R9M1	Q8R9M1 thermoaer
26	44	50.6	101	5	Q8IF11	Q8IF11 trypanosoma
27	44	50.6	175	10	Q8GSU7	Q8GSU7 coffea arab
28	44	50.6	177	16	Q9RKE7	Q9RKE7 streptomyce
29	44	50.6	297	10	Q8H527	Q8H527 oryza sativ
30	44	50.6	517	2	Q9ZGB6	Q9ZGB6 streptomyce
31	44	50.6	521	16	Q8ETN6	Q8ETN6 oceanobacil
32	44	50.6	553	16	Q9RUP7	Q9RUP7 deinococcus
33	44	50.6	1187	4	Q8IUP8	Q8IUP8 homo sapien
34	44	50.6	1413	13	Q9DER3	Q9DER3 xenopus lae
35	43	49.4	74	12	Q84349	Q84349 simian vtru
36	43	49.4	100	4	Q9UGS8	Q9UGS8 homo sapien
37	43	49.4	119	16	Q9KOM1	Q9KOM1 neisseria m
38	43	49.4	119	16	Q9JVP6	Q9JVP6 neisseria m
39	43	49.4	159	16	Q8RY14	Q8RY14 deinococcus
40	43	49.4	213	16	Q8NMT4	Q8NMT4 corynebacte
41	43	49.4	225	12	Q91NM4	Q91NM4 polyomaviru
42	43	49.4	225	12	Q91NN7	Q91NN7 polyomaviru
43	43	49.4	225	12	Q8IUB7	Q8IUB7 polyomaviru
44	43	49.4	225	12	Q910S1	Q910S1 polyomaviru
45	43	49.4	225	12	P90328	P90328 polyomaviru

## ALIGNMENTS

RESULT 1	Q8IV25	PRELIMINARY;	PRT;	329 AA.
ID	Q8IV25	PRELIMINARY;	PRT;	329 AA.
AC	Q8IV25;			
DT	01-MAR-2003 (TREMBLrel. 23, Created)			
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DE	Similar to C1q and tumor necrosis factor related protein 4.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_Taxid=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RA	Straussberg R.;			
RL	Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BC035628; AAH35628.1; -			
SQ	SEQUENCE 329 AA; 35256 MW; 16064DA8182A6732 CRC64;			
Query Match	100.0%; Score 87; DB 4; Length 329;			
Best Local Similarity	100.0%; Pred. No. 2.4e-05;			
Matches	16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1 MIPLLGLGPACWA 16			
DB	1 MIPLLGLGPACWA 16			
RESULT 2	Q8R066	PRELIMINARY;	PRT;	326 AA.
ID	Q8R066;			
AC	Q8R066;			
DT	01-JUN-2002 (TREMBLrel. 21, Created)			
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DE	Similar to C1q and tumor necrosis factor related protein 4.			
GN	0710001E1D1K.			
OS	Mus musculus (Mouse).			

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strauberg R.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC027315; AAH27315.1;  
 DR MGD: MGI:1914685; 0710001E10R1K.  
 DR InterPro: IPR001073; Clq.  
 DR Pfam: PF00386; Clq; 2.  
 DR SMART: SM00110; Clq; 2.  
 DR PROSITE: PS01113; Clq; 2.  
 SQ SEQUENCE 326 AA; 35057 MW; 7233917287B1051A CRC64;

Query Match 83.9%; Score 73; DB 11; Length 326;  
 Best Local Similarity 87.5%; Pred. No. 0.0032;  
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

DB 1 MLPLILGLGPACWA 16  
 1 MLPLILGLGPACWA 16

QY 1 MLPLILGLGPACWA 16  
 DB 1 MLPLILGLGPACWA 16

RESULT 3  
 ID 091YS9 PRELIMINARY; PRT; 993 AA.  
 AC 091YS9;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Similar to EphA3 (EC 2.7.1.112) (Ephrin receptor) (Tyrosine-protein kinase receptor).  
 GN EPHB3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strauberg R.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP + PROTEIN TYROSINE PHOSPHATE.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. EPHRIN RECEPTOR SUBFAMILY.  
 CC EMBL: BC014822; AAH14822.1;  
 DR MGD: MGI:104770; Ephb3.  
 DR InterPro: IPR001090; Ephrin\_receptor.  
 DR InterPro: IPR003962; Ptili\_Subd.  
 DR InterPro: IPR003961; Ptili\_III.  
 DR InterPro: IPR000719; Prot\_kinase.  
 DR InterPro: IPR001660; SAM.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR InterPro: IPR001426; YKase\_receptorV.  
 DR Pfam: PF01404; Eph\_Ibd. 1.  
 DR Pfam: PF00041; fn3\_2.  
 DR Pfam: PF00069; kinase. 1.  
 DR Pfam: PF00536; SAM; 1.  
 DR PRINTS: PR00014; FNTYPEIII.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR ProDom: PD001495; Ephrin\_receptor. 1.  
 DR ProDom: PD000001; Prot\_kinase. 1.  
 DR SMART: SM00615; Eph\_Ibd. 1.  
 DR SMART: SM00060; fn3\_2.  
 DR SMART: SM00454; SAM; 1.  
 DR SMART: SM00219; TyrcK. 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP. 1.  
 DR PROSITE: PS00011; PROTEIN\_KINASE\_DOM. 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR. 1.  
 DR PROSITE: PS00790; RECEPTOR\_TYR\_KIN\_V\_1; 1.  
 DR PROSITE: PS00791; RECEPTOR\_TYR\_KIN\_V\_2; 1.

DR PROSITE: PS00105; SAM\_DOMAIN. 1.  
 KW ATP-binding; Glycoprotein; Kinase; Phosphorylation; Receptor; Repeat;  
 KW Transferase; Transmembrane; Tyrosine-Protein kinase.  
 SQ SEQUENCE 993 AA; 109661 MW; 0706E2AC98A81B4 CRC64;

Query Match 65.5%; Score 57; DB 11; Length 993;  
 Best Local Similarity 68.8%; Pred. No. 2.4;  
 Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MLPLILGLGPACWA 16  
 DB 14 LAPLILPLILPACWA 29

QY 1 MLPLILGLGPACWA 16  
 DB 14 LAPLILPLILPACWA 29

RESULT 4  
 ID 051537 PRELIMINARY; PRT; 115 AA.  
 AC 051537;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Type 4 fimbrial biogenesis protein PILY2.  
 GN PILY2 OR PA435.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=95157261; PubMed=7854110;  
 RA Russell M.A., Darzi A.;  
 RT "The pili gene product of Pseudomonas aeruginosa, required for pili biogenesis, shares amino acid sequence identity with the N-termini of type 4 prepilin proteins.";  
 RT Mol. Microbiol. 13:973-985 (1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=96037793; PubMed=7565109;  
 RA Alm R.A., Mattick J.S.;  
 RT "Identification of a gene, pilV, required for type 4 fimbrial biogenesis in Pseudomonas aeruginosa, whose product possesses a pre-pilin-like leader sequence.";  
 RT Mol. Microbiol. 16:485-496 (1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RA Alm R.A., Hallinan J.P., Watson A.N., Mattick J.S.;  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Micozuchi S.D., Warren P., Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M., Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y., Brody L.L., Coulter S.N., Folger K.R., Kae A., Larbig K., Lim R.M., Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T., Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.";  
 RT Nature 406:959-964 (2000).  
 RL Nature 406:959-964 (2000).  
 DR EMBL: L76605; AAA93503.1;  
 DR EMBL: AB004869; AAC07943.1;  
 DR Complete proteome.  
 SQ SEQUENCE 115 AA; 12986 MW; AC19C31640473181 CRC64;

Query Match 59.8%; Score 52; DB 16; Length 115;  
 Best Local Similarity 56.2%; Pred. No. 1.9;  
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MLPLILGLGPACWA 16

Db 3 VLPELILALVPGICLWA 18

RESULT 5

Q8PCRO PRELIMINARY; PRT; 155 AA.

AC Q8PCRO; 01-OCT-2002 (TREMBlrel. 22, Created)

DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)

DE 01-OCT-2002 (TREMBlrel. 22, Last annotation update)

DE Hypoetical protein XCC0653.

GN XCC0653.

OS Xanthomonas campestris (pv. campestris).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

OC Xanthomonadaceae; Xanthomonas.

OX NCBI\_TaxID=340;

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 33913 / NCPPB 528;

RX MEDLINE=22022145; PubMed=12024217;

RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,

RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,

RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,

RA Camarotte G., Camavan F., Cardozo J., Chambergo F., Clapina L.P.,

RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,

RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,

RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,

RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,

RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,

RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,

RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,

RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,

RA Spindola L.A.F., Takita M.A., Tamura R.B., Teixeira E.C., Tezza R.I.D.,

RA Trindade dos Santos M., Truffi D., Teat S.M., White F.F.,

RA Setubal J.C., Kitajima J.P.

RT "Comparison of the genomes of two Xanthomonas pathogens with differing

RT host specificities."

RL Nature 417:459-463(2002).

DR EMBL; AE012163; AM39969.1;

KW Hypoetical protein; Complete proteome.

SQ SEQUENCE 155 AA; 17448 MW; DE92F2FA2C63A51 CRC64;

Query Match 56.3%; Score 49; DB 16; Length 155;

Best Local Similarity 60.0%; Pred. No. 7.1;

Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 MLEPLILGLGPACW 15

Db 91 MRPWRLSLTGPQACW 105

RESULT 6

Q8C883 PRELIMINARY; PRT; 998 AA.

AC Q8C883; 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE Protocadherin 12.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Head;

RX MEDLINE=22354683; PubMed=12466851;

RA The FANTOM Consortium;

RA The RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs."

RL Nature 420:563-573(2002).

DR EMBL; AK048100; BAC33242.1; -

SQ SEQUENCE 998 AA; 109373 MW; 604080E3D5C37F78 CRC64;

Query Match 56.3%; Score 49; DB 11; Length 998;

Best Local Similarity 69.2%; Pred. No. 39;

Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MLEPLILGLGPAA 13

Db 4 LLEPLILGLGPAS 16

RESULT 7

Q9V9L4 PRELIMINARY; PRT; 1061 AA.

AC Q9V9L4; 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE CG11212 protein.

GN CG11212.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abdiril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolhakov S.,

RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,

RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson R., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Fostin C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodok A., Gong F., Gorell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Modarres C., Morris J., Mostrel A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,

RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirydas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Messarman D.A., Weinstein G.M., Weisenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu C., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,

RT "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195(2000).

DR EMBL; AE003784; AAF57274.1; -

DR FlyBase; FBgn0033068; CG11212.

DR InterPro; IPR003392; Patched.

DR InterPro; IPR006162; Ppatnme\_attach.

DR InterPro; IPR000731; SSD\_5TM.

DR Pfam; PF02460; Patched; 1.  
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.  
 DR PROSITE; PS0156; SSD; 1.  
 SQ SEQUENCE 1061 AA; 120880 MW; D6FF3237BBB9A96 CRC64;

Query Match 56.3%; Score 49; DB 5; Length 1061;  
 Best Local Similarity 57.1%; Pred. No. 41;  
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 Db 1 MLPLGLGAPAC 14  
 740 LPLPLSLFPGSC 753

## RESULT 8

ID Q9DCM1 PRELIMINARY; PRT; 379 AA.  
 AC Q9DCM1;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE 061000922Rik protein.  
 GN 061000922Rik.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Kidney;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinaawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 Schirml L.M., Staulli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 Sakai K., Okido T., Furuno M., Iono H., Baldarelli R., Barab G.,  
 Blake J., Botfield D., Boujane N., Carmignoli P., de Bonaldo M.F.,  
 Brownstein W.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 Garslind S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,  
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kontseki S.,  
 Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.  
 DR ENBL; AK002414; BAB22082.1; -.  
 DR HSP; P28990; ICHC.  
 DR MGD; MGI:1918923; 061000922Rik.  
 DR InterPro; IPR001841; Znf\_ring.  
 DR Pfam; PF00097; zf-C3HC4\_1.  
 DR SMART; SM00184; RING\_1.  
 DR PROSITE; PS50089; ZF\_RING\_2; 1.  
 KW Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 379 AA; 41577 MW; DA36AA22623B3901 CRC64;

Query Match 55.2%; Score 48; DB 11; Length 379;  
 Best Local Similarity 59.2%; Pred. No. 23;  
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MLPLGLGAPAC 13  
 31 LPLPLSLFPGSC 43

RESULT 9  
 Q94RE8

ID Q94RE8 PRELIMINARY; PRT; 569 AA.  
 AC Q94RE8;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE NADH dehydrogenase subunit 5 (Fragment).  
 GN NDS.  
 OS Lithobius forficatus.  
 OC Eukaryota; Metazoa; Arthropoda; Myriapoda; Chilopoda;  
 OC Pleurostigmophora; Lithobionomorpha; Lithobiidae; Lithobius.  
 NC NCB1\_TaxID=7552;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Huang U.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21441907; PubMed=11557978;  
 RA Huang U., Friedrich M., Choe C., Kim W.;  
 RT "Mitochondrial Protein Phylogeny joins myriapods with chelicerates."  
 RL Nature 413:154-157(2001).  
 DR ENBL; A0270997; CAC6943.1; -.  
 DR InterPro; IPR003916; NADHbd\_oxred5.  
 DR InterPro; IPR001750; Oxidored\_q1.  
 DR InterPro; IPR001516; Oxidored\_q1\_N.  
 DR InterPro; IPR006162; Ppantne\_attach.  
 DR Pfam; PF00361; oxidored\_q1\_1.  
 DR Pfam; PF00662; oxidored\_q1\_N\_1.  
 DR PRINTS; PR01434; NADHbGNASE5.  
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.  
 KW NAD; Oxidoreductase; Ubiquinone; Mitochondrion.  
 FT NON TER 569 569  
 SQ SEQUENCE 569 AA; 62591 MW; 4C3104BFA6F72AB CRC64;

Query Match 54.0%; Score 47; DB 8; Length 569;  
 Best Local Similarity 58.3%; Pred. No. 47;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 4 LPLGLGAPACW 15  
 169 LPLGLGAPACW 180

## RESULT 10

ID Q8P3B9 PRELIMINARY; PRT; 719 AA.  
 AC Q8P3B9;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DE Transglycolase.  
 GN XCC4157.  
 OS Xanthomonas campestris (pv. campestris).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.  
 NC NCB1\_TaxID=340;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33913 / NCPPB 528;  
 RX MEDLINE=22022145; PubMed=12024217;  
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
 Queglio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
 Catarote G., Canavan F., Cardoso J., Chambergio F., Ciapina L.P.,  
 Clarel R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,  
 Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
 Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,

RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira F.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
 RA Setubal J.C., Kitajima J.P.,  
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
 RT host specificities.";  
 RL Nature 417:459-463(2002).  
 DR EMBL: AE012541; AAM43373.1; -;  
 DR InterPro: IPR001823; Ald1\_epimerase.  
 DR Pfam: PF01263; Aldose\_epim; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 719 AA; 77224 MW; 45F5583D9E83D132 CRC64;

Query Match 53.4%; Score 46.5; DB 16; Length 719;  
 Best Local Similarity 73.3%; Pred. No. 70;  
 Matches 11; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 2 LPIILGLGPACMA 16  
 Db 8 LPIILGL-LAPFAAMA 21

RESULT 11  
 O9AK63 PRELIMINARY; PRT; 228 AA.

AC O9AK63;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Putative integral membrane protein.  
 GN SC04054 OR 28CD60.20.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycetales; Streptomycetaceae; Streptomyces.  
 RX NCBI\_TaxID=1902;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Croftin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neil S.,  
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wierozurek A., Woodward J., Barrett B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL: AL939118; CAC32325.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 228 AA; 25240 MW; C13D0696B40D2991 CRC64;

Query Match 52.9%; Score 46; DB 16; Length 228;  
 Best Local Similarity 60.0%; Pred. No. 29;  
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 MLPILGLGPACW 15  
 Db 1 MLPILGLASAAACW 15

RESULT 12  
 O967F2 PRELIMINARY; PRT; 496 AA.

AC O967F2;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Hypothetical 57.2 kDa protein (Putative alpha 2  
 DE mannosyltransferase).  
 GN T27F7.3 OR PG-B.

OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA Waterston R.;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Halleworth K.;  
 RT "The sequence of C. elegans cosmid T27F7";  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Oriol R., Martinez-Duncker I., Chantret I., Mollicone R., Codogno P.;  
 RT "Common origin and evolution of glycosyltransferases using Dol-P-  
 RT monosaccharides as donor substrate.";  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U58762; AAK9304.2; -;  
 DR EMBL: A0431373; CAD24083.1; -;  
 DR WormBep: T27F7.3a; CE30458.  
 DR InterPro: IPR005599; PMP.  
 DR Pfam: PF03901; PMP; 1.  
 KW Hypothetical protein: Glycosyltransferase; Transferase.  
 SQ SEQUENCE 496 AA; 57191 MW; 6A0C53B5346D78C CRC64;

Query Match 52.9%; Score 46; DB 5; Length 496;  
 Best Local Similarity 63.6%; Pred. No. 59;  
 Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLPILGLGP 11  
 Db 270 MLPILGLGP 280

RESULT 13  
 O8CHQ8 PRELIMINARY; PRT; 517 AA.

AC O8CHQ8;  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Similar to N-acetylglucosamine-1-phosphodiester  
 DE alpha-N-acetylglucosaminidase.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N;  
 RA Strausberg R.;  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC039790; AAH39790.1; -;  
 SQ SEQUENCE 517 AA; 56013 MW; 80FEE9AFB3873177 CRC64;

Query Match 52.9%; Score 46; DB 11; Length 517;  
 Best Local Similarity 64.3%; Pred. No. 61;  
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MLPILGLGPAC 14  
 Db 1 MLPILGLGPAC 14

Db 11 LIPALLGLGVAMC 24

## RESULT 14

Q8B348 PRELIMINARY; PRT; 517 AA.  
 AC Q8B348;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE N-acetylglucosamine-1-PHOSPHODIESTER  
 DE alpha-N-acetylglucosaminidase.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573 (2002).  
 DR EMBL; AK032158; BAC27731.1;  
 SQ SEQUENCE 517 AA; 56056 MW; 177F813D20552C4B CRC64;

## Query Match

Best Local Similarity 52.9%; Score 46; DB 11; Length 517;  
 Pred. No. 61;  
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MLPLLGLGSPAC 14  
 Db 11 LIPALLGLGVAMC 24

## RESULT 15

Q9UVMS PRELIMINARY; PRT; 558 AA.  
 AC Q9UVMS;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Mea1.2.  
 GN MFS1.2.  
 OS Coprinus cinereus (inky cap fungus).  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
 OC Agaricales; Psathyrellaceae; Coprinopsis.  
 OX NCBI\_TaxId=5346;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JV6;  
 RX MEDLINE=20221720; PubMed=10757757;  
 RA Halsall J.R., Milner M.J., Casselton L.A.;  
 RT "Three Subfamilies of Pheromone and Receptor Genes Generate Multiple B  
 RT Mating Specificities in the Mushroom Coprinus cinereus."  
 RL Genetics 154:1115-1123 (2000).  
 CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL; AF186392; AA01427.1; -.  
 DR InterPro; IPR007114; MFS.  
 DR InterPro; IPR005828; Sub\_transporter.  
 DR Pfam; PF00063; sugar\_tr\_1.  
 DR PRINTS; PR01036; TCRTRTB.  
 DR PROSITE; PS50850; MFS; 1.  
 DR Transmembrane.  
 SQ SEQUENCE 558 AA; 59446 MW; 6F3232638C26A2AF CRC64;

Query Match 52.9%; Score 46; DB 3; Length 558;  
 Best Local Similarity 53.3%; Pred. No. 66;  
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 MLPLLGLGSPACW 15  
 Db 255 LVPLLGLGLTAMW 269

Search completed: January 12, 2004, 08:19:14  
 Job time: 7.20359 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:09 ; Search time 2.4118 Seconds  
(without alignments)  
1053.272 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_1\_16

Perfect score: 87

Sequence: 1 MLPILGLGPACWA 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_19Jun03:\*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*

2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*

3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*

4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*

5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*

6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*

7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*

8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*

9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*

10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*

11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*

12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*

13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*

14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*

15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*

16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*

17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*

18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*

19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*

20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*

21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	100.0	17	22	AAB61467
2	87	100.0	133	23	ABG70386
3	87	100.0	284	23	ABG70384
4	87	100.0	299	23	ABG70385
5	87	100.0	329	22	AAB61424
6	87	100.0	329	22	AAB61606
7	87	100.0	329	23	ABG79643
8	87	100.0	348	22	AAB61423
9	73	83.9	199	22	AAB61488

10	72	82.8	14	22	AAB61465
11	52	59.8	115	21	AAV52384
12	49	56.3	1061	22	ABE65553
13	49	56.3	1180	19	AAB61273
14	48	55.2	912	22	AAU52873
15	47	54.0	167	22	ABG17123
16	47	54.0	516	23	ABO6122
17	47	54.0	656	19	AAW71648
18	45.5	52.3	101	22	AAO10317
19	45.5	52.3	120	22	AAO10647
20	45	51.7	16	21	AAV93254
21	45	51.7	18	19	AAW71681
22	45	51.7	18	22	AAV98994
23	45	51.7	347	21	ABR16457
24	45	51.7	378	21	ABR42695
25	45	51.7	378	24	ABU07370
26	45	51.7	602	22	ABR71195
27	44	50.6	54	22	ABG08743
28	44	50.6	90	22	AAW84788
29	44	50.6	318	17	AAW88105
30	44	50.6	318	18	AAW18343
31	44	50.6	390	22	AAV92502
32	44	50.6	1184	20	AAV41750
33	44	50.6	1184	21	AAW44306
34	44	50.6	1184	22	AAU12368
35	44	50.6	1184	24	ABU67042
36	44	50.6	1184	24	ABU59847
37	44	50.6	1184	24	ABU61136
38	44	50.6	1184	24	ABU03542
39	44	50.6	1187	22	AAE06028
40	44	50.6	1187	22	AAE06584
41	44	50.6	76	22	AAW06512
42	43	49.4	96	17	AAW00723
43	43	49.4	96	17	AAW11890
44	43	49.4	96	17	AAW11890
45	43	49.4	119	21	AAV75262

#### ALIGNMENTS

RESULT 1

AAB61467

ID AAB61467 standard; peptide; 17 AA.

XX AAB61467;

AC

XX

XX

DT 04-APR-2001 (first entry)

XX

XX

DE MANGO 245 signal peptide #2.

XX

XX

KW TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;

KW autoimmune; allergy; cardiovascular; brain; degenerative; placental;

KW pancreatic; skeletal; muscle.

XX

XX

OS Catarrhini sp..

XX

XX

PN WO200100672-A1.

XX

XX

PD 04-JAN-2001.

XX

XX

PF 29-JUN-2000; 2000WO-US18184.

XX

XX

PR 29-JUN-1999; 99US-0342687.

XX

XX

(MILL-) MILLENNIUM PHARM INC.

PA

XX

PI Holtzman DA, Barnes TM, Frazer CC, Sharp JD;

XX

XX

DR WPI; 2001-050127/06.

XX

XX

PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and

PT MANGO 245 proteins, useful in the treatment of inflammatory diseases

MANGO 245 signal p

Pseudomonas aerugi

Drosophila melanog

Homo sapiens proto

Propionibacterium

Novel human diagno

Human NS protein s

Fluorescent calmod

Human polypeptide

Human polypeptide

Signal peptide use

Endoplasmic reticu

N-terminal calreti

Human nervous syst

Human ORFX ORF245

Human ORFX NOV2

Human protein NOV2

Drosophila melanog

Novel human diagno

Human immune/haema

Bovine p32 11-cis-

Bovine p32 11-cis-

Human protein sequ

Human PRO731 prote

Human PRO731 (UNQ3

Human PRO731 polyp

Human PRO polypept

Human secreted/cra

Novel secreted and

Human PRO731 polyp

Angiogenesis-assoc

Human asthma-assoc

Human protein sequ

Human foetal prote

Murine fibrosin 2B

Murine fibrosin 2B

Neisseria gonorrhoe

PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
 disorders (e.g. jaundice) -  
 PS Disclosure; Page 243; 262pp; English.  
 XX  
 CC The present invention relates to cDNAs encoding TANGO 244,  
 CC TANGO 246, TANGO 275, TANGO 300 and TANGO 245 proteins.  
 CC The nucleic acids, proteins and protein modulators are useful for  
 CC treating colonic disorders, inflammatory diseases, tumors,  
 CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
 CC allergic diseases, cardiovascular diseases, brain disorders,  
 CC degenerative diseases placental, pancreatic, skeletal and muscle  
 CC disorders.  
 XX  
 CC  
 XX Sequence 17 AA;  
 SQ  
 Query Match 100.0%; Score 87; DB 22; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-07;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 MPELLGLGPPACWA 16  
 Db 1 MPELLGLGPPACWA 16  
 RESULT 2  
 ABG70386  
 ID ABG70386 standard; Protein; 133 AA.  
 XX  
 AC ABG70386;  
 XX  
 DT 05-NOV-2002 (first entry)  
 DE Adipocyte complement-related C1q Tumour Necrosis Factor-like protein #4.  
 XX  
 KW Human; NOX; NOX-associated disorder; cardiomyopathy; atherosclerosis;  
 KW cell signal processing; metabolic pathway modulation; metabolic disorder;  
 KW obesity; diabetes; infectious disease; neurodegenerative disorder; acne;  
 KW Alzheimer's disease; Parkinson's disease; immune disorder; cancer;  
 KW haematopoietic disorder; cirrhosis; pancreatitis; learning defect;  
 KW memory defect; infertility; congenital heart defect; hair growth;  
 KW pigmentation disorder; endocrine disorder; respiratory disease; health;  
 KW gastro-intestinal disease; reproductive; neurological disease;  
 KW bone marrow transplantation; endocrine disease; allergy; inflammation;  
 KW nephrological disorder; urinary system disorder; age-related disorder;  
 KW neuropsychiatric disorder; BGF-related protein; SCUBE1; TEN-M4;  
 KW adipocyte complement-related C1q tumour necrosis factor; out at first;  
 KW beta adrenergic receptor kinase; EphA6/enk-2; glucose transporter;  
 KW type 1a membrane sushi-containing domain; butyrophilin;  
 KW type 1a membrane-sushi domain containing.  
 KW  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200257453-A2.  
 XX  
 PD 25-JUL-2002.  
 XX  
 PF 19-DEC-2001; 2001WO-US50331.  
 XX  
 PF 19-DEC-2000; 2000US-265704P.  
 PR 20-DEC-2000; 2000US-257314P.  
 PR 02-MAY-2001; 2001US-288153P.  
 PR 29-MAY-2001; 2001US-294075P.  
 PR 24-JUL-2001; 2001US-307506P.  
 PR 10-AUG-2001; 2001US-311590P.  
 PR 10-AUG-2001; 2001US-311613P.  
 PR 29-AUG-2001; 2001US-315617P.  
 PR 14-SEP-2001; 2001US-322358P.  
 PR  
 XX  
 XX (CURA-) CURAGEN CORP.  
 PI Gangolli EA, Paturnjan M, Vernet CM, Malyankar UM, Kekuda R;  
 PI Stone DJ, Anderson D, Shimkets RA, Burgess CE, Zernusen BD, Liu X;

PI Spytek KA, Casman ST, Boldog FL, Smithson G, Li L, Ji W;  
 XX WPI; 2002-590744/63.  
 DR N-PSDB; ABS52098.  
 XX  
 PT Novel isolated NOX polypeptide useful for treating cardiomyopathy,  
 PT atherosclerosis, metabolic disorders, diabetes, obesity, infectious  
 PT disease, anorexia, neurodegenerative disorders, Alzheimer's disease or  
 PT cancer -  
 PS Claim 1; Page 32; 318pp; English.  
 XX  
 CC The present invention relates to new NOX polypeptides. The invention is  
 CC useful for treating or preventing a NOX-associated disorder such as  
 CC cardiomyopathy or atherosclerosis, where the disorder is related to cell  
 CC signal processing and metabolic pathway modulation in a subject.  
 CC preferably human. The invention is also useful for treating metabolic  
 CC disorders (e.g. obesity), diabetes, infectious disease, neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease), immune  
 CC disorders, haematopoietic disorders and various cancers. The molecules of  
 CC the invention are also useful for treating or preventing cirrhosis,  
 CC pancreatitis, learning and memory defects, infertility, congenital heart  
 CC defects, acne, hair growth, pigmentation disorders, endocrine disorder,  
 CC respiratory disease, gastro-intestinal diseases, reproductive, health,  
 CC neurological diseases, bone marrow transplantation, endocrine diseases,  
 CC allergy and inflammation, nephrological disorders, urinary system  
 CC disorders, neuropsychiatric disorders and age-related disorders.  
 CC The present amino acid sequence represents a NOX protein of the  
 CC invention.  
 XX  
 SQ Sequence 133 AA;  
 Query Match 100.0%; Score 87; DB 23; Length 133;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 MPELLGLGPPACWA 16  
 Db 1 MPELLGLGPPACWA 16  
 RESULT 3  
 ABG70384  
 ID ABG70384 standard; Protein; 284 AA.  
 XX  
 AC ABG70384;  
 XX  
 DT 05-NOV-2002 (first entry)  
 DE Adipocyte complement-related C1q Tumour Necrosis Factor-like protein #2.  
 XX  
 KW Human; NOX; NOX-associated disorder; cardiomyopathy; atherosclerosis;  
 KW cell signal processing; metabolic pathway modulation; metabolic disorder;  
 KW obesity; diabetes; infectious disease; neurodegenerative disorder; acne;  
 KW Alzheimer's disease; Parkinson's disease; immune disorder; cancer;  
 KW haematopoietic disorder; cirrhosis; pancreatitis; learning defect;  
 KW memory defect; infertility; congenital heart defect; hair growth;  
 KW pigmentation disorder; endocrine disorder; respiratory disease; health;  
 KW gastro-intestinal disease; reproductive; neurological disease;  
 KW bone marrow transplantation; endocrine disease; allergy; inflammation;  
 KW nephrological disorder; urinary system disorder; age-related disorder;  
 KW neuropsychiatric disorder; BGF-related protein; SCUBE1; TEN-M4;  
 KW adipocyte complement-related C1q tumour necrosis factor; out at first;  
 KW beta adrenergic receptor kinase; EphA6/enk-2; glucose transporter;  
 KW type 1a membrane sushi-containing domain; butyrophilin;  
 KW type 1a membrane-sushi domain containing.  
 KW  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200257453-A2.  
 XX  
 PD 25-JUL-2002.  
 XX



PF 19-DEC-2001; 2001WO-US50331.  
XX 19-DEC-2000; 2000US-265704P.  
PR 20-DEC-2000; 2000US-257314P.  
PR 02-MAY-2001; 2001US-288153P.  
PR 29-MAY-2001; 2001US-294075P.  
PR 24-JUL-2001; 2001US-307506P.  
PR 10-AUG-2001; 2001US-311590P.  
PR 10-AUG-2001; 2001US-311613P.  
PR 29-AUG-2001; 2001US-315617P.  
PR 14-SEP-2001; 2001US-322358P.  
XX  
XX (CURA-) CURAGEN CORP.  
PI Gangolli EA, Patturajan M, Vernet CAM, Malyanar UM, Kekuda R,  
PI Stone DJ, Anderson D, Shimkets RA, Burgess CE, Zehnsen BD, Liu X,  
PI Spytek KA, Casman SU, Boldog FL, Smithson G, Li L, Ji W,  
XX WPI; 2002-590744/63.  
XX N-PSDB; ABS52096.  
XX  
XX Novel isolated NOVX polypeptide useful for treating cardiomyopathy,  
PT atherosclerosis, metabolic disorders, diabetes, obesity, infectious  
PT disease, anorexia, neurodegenerative disorders, Alzheimer's disease or  
PT cancer  
XX  
XX Claim 1; Page 30; 318pp; English.  
XX  
XX The present invention relates to new NOVX polypeptides. The invention is  
CC useful for treating or preventing a NOVX-associated disorder such as  
CC cardiomyopathy or atherosclerosis, where the disorder is related to cell  
CC signal processing or metabolic pathway modulation in a subject.  
CC preferably human. The invention is also useful for treating metabolic  
CC disorders (e.g. obesity), diabetes, infectious disease, neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), immune  
CC disorders, haematopoietic disorders and various cancers. The molecules of  
CC the invention are also useful for treating or preventing cirrhosis,  
CC pancreatitis, learning and memory defects, infertility, congenital heart  
CC defects, acne, hair growth, pigmentation disorders, endocrine disorders,  
CC respiratory diseases, gastro-intestinal diseases, reproductive, health,  
CC neurological diseases, bone marrow transplantation, endocrine diseases,  
CC allergy and inflammation, nephrological disorders, urinary system  
CC disorders, neuropsychiatric disorders and age-related disorders.  
CC The present amino acid sequence represents a NOVX protein of the  
CC invention.  
XX  
XX Sequence 284 AA;  
SQ  
Query Match 100.0%; Score 87; DB 23; Length 284;  
Best Local Similarity 100.0%; Pred. No. 6.7e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 MLPLILGLGPAACWA 16  
DB 1 MLPLILGLGPAACWA 16  
RESULT 4  
ABG70385  
ID ABG70385 standard; Protein; 299 AA.  
XX  
XX ABG70385;  
AC  
XX  
XX 05-NOV-2002 (first entry)  
DT  
XX  
XX Adipocyte complement-related C1q Tumour Necrosis Factor-like protein #3.  
DE  
XX  
XX Human; NOVX; NOVX-associated disorder; cardiomyopathy; atherosclerosis;  
KM cell signal processing; metabolic pathway modulation; metabolic disorder;  
KM obesity; diabetes; infectious disease; neurodegenerative disorder; acne;  
KM Alzheimer's disease; Parkinson's disease; immune disorder; cancer;  
KM haematopoietic disorder; cirrhosis; pancreatitis; learning defect;  
KM memory defect; infertility; congenital heart defect; hair growth;

KW pigmentation disorder; endocrine disorder; respiratory disease; health;  
KM gastro-intestinal disease; reproductive; neurological disease;  
KM bone marrow transplantation; endocrine disease; allergy; inflammation;  
KM nephrological disorder; urinary system disorder; age-related disorder;  
KM neuropsychiatric disorder; EGF-related protein; SCUB1; TEN-M4;  
KM adipocyte complement-related C1q tumour necrosis factor; out at first;  
KM beta adrenergic receptor kinase; EphA6/ehk-2; glucose transporter;  
KM type 1a membrane-sushi domain containing.  
XX  
XX Homo sapiens.  
XX  
XX WO200257453-A2.  
XX  
XX 25-JUL-2002.  
XX  
XX 19-DEC-2001; 2001WO-US50331.  
XX  
XX 19-DEC-2000; 2000US-265704P.  
PR 20-DEC-2000; 2000US-257314P.  
PR 02-MAY-2001; 2001US-288153P.  
PR 29-MAY-2001; 2001US-294075P.  
PR 24-JUL-2001; 2001US-307506P.  
PR 10-AUG-2001; 2001US-311590P.  
PR 10-AUG-2001; 2001US-311613P.  
PR 29-AUG-2001; 2001US-315617P.  
PR 14-SEP-2001; 2001US-322358P.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Gangolli EA, Patturajan M, Vernet CAM, Malyanar UM, Kekuda R,  
XX Stone DJ, Anderson D, Shimkets RA, Burgess CE, Zehnsen BD, Liu X,  
XX Spytek KA, Casman SU, Boldog FL, Smithson G, Li L, Ji W,  
XX WPI; 2002-590744/63.  
XX N-PSDB; ABS52097.  
XX  
XX Novel isolated NOVX polypeptide useful for treating cardiomyopathy,  
PT atherosclerosis, metabolic disorders, diabetes, obesity, infectious  
PT disease, anorexia, neurodegenerative disorders, Alzheimer's disease or  
PT cancer  
XX  
XX Claim 1; Page 31; 318pp; English.  
XX  
XX The present invention relates to new NOVX polypeptides. The invention is  
CC useful for treating or preventing a NOVX-associated disorder such as  
CC cardiomyopathy or atherosclerosis, where the disorder is related to cell  
CC signal processing or metabolic pathway modulation in a subject.  
CC preferably human. The invention is also useful for treating metabolic  
CC disorders (e.g. obesity), diabetes, infectious disease, neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), immune  
CC disorders, haematopoietic disorders and various cancers. The molecules of  
CC the invention are also useful for treating or preventing cirrhosis,  
CC pancreatitis, learning and memory defects, infertility, congenital heart  
CC defects, acne, hair growth, pigmentation disorders, endocrine disorders,  
CC respiratory diseases, gastro-intestinal diseases, reproductive, health,  
CC neurological diseases, bone marrow transplantation, endocrine diseases,  
CC allergy and inflammation, nephrological disorders, urinary system  
CC disorders, neuropsychiatric disorders and age-related disorders.  
CC The present amino acid sequence represents a NOVX protein of the  
CC invention.  
XX  
XX Sequence 299 AA;  
SQ  
Query Match 100.0%; Score 87; DB 23; Length 299;  
Best Local Similarity 100.0%; Pred. No. 7e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 MLPLILGLGPAACWA 16  
DB 1 MLPLILGLGPAACWA 16



PN WO200262841-A2.  
 XX  
 PD 15-AUG-2002.  
 XX  
 PF 28-JAN-2002; 2002WO-US02616.  
 XX  
 PR 02-FEB-2001; 2001US-266195P.  
 PR 08-FEB-2001; 2001US-267924P.  
 PR 09-FEB-2001; 2001US-267816P.  
 PR 09-FEB-2001; 2001US-268112P.  
 PR 26-FEB-2001; 2001US-271639P.  
 PR 07-SEP-2001; 2001US-317818P.  
 PR 21-DEC-2001; 2001US-343553P.  
 XX  
 PA (INCYTE GENOMICS INC.  
 XX  
 PI Tang TY, Yue H, Gandhi AR, Yao MG, Warren BA, Ding L, Duggan BM;  
 PI Xu Y, Yang J, Thangavelu K, Lal PG, Honchell CD, Walla NK, Lee S;  
 PI Lee EA, Richardson TW, Baughn MR, Elliott VS;  
 XX  
 DR WPI: 2002-657522/70.  
 DR N-PSDB; ABS64954.  
 XX  
 PT New human secreted proteins and nucleic acids useful in diagnosing,  
 PT treating and preventing cell proliferative, autoimmune/inflammatory,  
 PT cardiovascular, neurological, and developmental disorders -  
 PT  
 PS Claim 1; Page 140; 158pp; English.  
 XX  
 CC The invention relates to twenty four human secreted proteins  
 CC (SECP1-24), proteins 90% identical to them and active fragments of them.  
 CC Also included are nucleic acids encoding the SECP proteins, a recombinant  
 CC polynucleotide comprising a promoter sequence operably linked to the  
 CC nucleic acid, a cell transformed with the recombinant polynucleotide,  
 CC a transgenic organism comprising the recombinant polynucleotide, an  
 CC anti-SECP antibody, and screening for ant/agonists and modulators of  
 CC SECP function or expression. The SECP proteins and nucleic acids are  
 CC useful in the diagnosis, treatment and prevention of cell proliferative  
 CC (e.g. actinic keratosis, arteriosclerosis, burns, hepatitis or  
 CC cancer), autoimmune/inflammatory (e.g. AIDS (acquired immunodeficiency  
 CC syndrome), asthma, anaemia, allergies or atopic dermatitis),  
 CC cardiovascular (e.g. congestive heart failure, ischemic heart disease,  
 CC myocardial infarction, hypertensive heart disease, or vascular tumors),  
 CC neurological (e.g. epilepsy, stroke, cerebral neoplasms, or Alzheimer's  
 CC disease), and developmental (e.g. renal tubular acidosis, Cushing's  
 CC syndrome, Duchenne and Becker muscular dystrophy, or hypothyroidism)  
 CC disorders. Many other diseases and disorders are listed in the  
 CC specification. These may also be used in assessing the effects of  
 CC exogenous compounds on the expression of nucleic acid and amino acid  
 CC sequences of the secreted proteins. The present sequence represents a  
 CC SECP protein of the invention.  
 CC  
 SQ Sequence 329 AA;  
 XX  
 XX  
 Query Match 100.0%; Score 87; DB 23; Length 329;  
 Best Local Similarity 100.0%; Pred. No. 7.8e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 MLPLLLGLGPAACWA 16  
 |||||  
 DB 1 MLPLLLGLGPAACWA 16  
 |||||  
 RESULT 8  
 ID AAB61423 standard; protein; 348 AA.  
 XX  
 AC AAB61423;  
 XX  
 DT 04-APR-2001 (first entry)  
 XX  
 DE Human MANGO 245 protein.  
 XX

KM TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
 KM autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
 KM pancreatic; skeletal; muscle.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200100672-A1.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 29-JUN-2000; 2000WO-US18184.  
 XX  
 PR 29-JUN-1999; 99US-0342687.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Holtzman DA, Barnes TW, Frazer CC, Sharp JD;  
 PI  
 DR WPI: 2001-050127/06.  
 XX  
 PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
 PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
 PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
 PT disorders (e.g. jaundice) -  
 PT  
 PS Claim 1; Fig 23; 262pp; English.  
 XX  
 CC The present invention relates to cDNAs encoding TANGO 244,  
 CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
 CC The nucleic acids, proteins and protein modulators are useful for  
 CC treating colonic disorders, inflammatory diseases, tumors,  
 CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
 CC allergic diseases, cardiovascular diseases, brain disorders,  
 CC degenerative diseases placental, pancreatic, skeletal and muscle  
 CC disorders.  
 CC  
 SQ Sequence 348 AA;  
 XX  
 XX  
 Query Match 100.0%; Score 87; DB 22; Length 348;  
 Best Local Similarity 100.0%; Pred. No. 8.3e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 MLPLLLGLGPAACWA 16  
 |||||  
 DB 1 MLPLLLGLGPAACWA 16  
 |||||  
 RESULT 9  
 ID AAB61488 standard; protein; 199 AA.  
 XX  
 AC AAB61488;  
 XX  
 DT 04-APR-2001 (first entry)  
 XX  
 DE Murine MANGO 245 protein.  
 XX  
 KM TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
 KM autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
 KM pancreatic; skeletal; muscle.  
 XX  
 OS Mus musculus.  
 XX  
 PN WO200100672-A1.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 29-JUN-2000; 2000WO-US18184.  
 XX  
 PR 29-JUN-1999; 99US-0342687.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX

PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;  
XX WPI, 2001-050127/06.  
XX  
XX Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
XX disorders (e.g. jaundice) -  
XX  
XX Claim 1; Fig 29; 262pp; English.  
XX  
XX The present invention relates to cDNAs encoding TANGO 244,  
CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
CC The nucleic acids, proteins and protein modulators are useful for  
CC treating colonic disorders, inflammatory diseases, tumors,  
CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
CC allergic diseases, cardiovascular diseases, brain disorders,  
CC degenerative diseases placental, pancreatic, skeletal and muscle  
CC disorders.  
XX  
XX Sequence 199 AA;  
SQ  
XX  
XX Query Match 83.9%; Score 73; DB 22; Length 199;  
Best Local Similarity 87.5%; Pred. No. 0.00082;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 1 MFLPLGLGPPAAC 16  
1 ||||| |||||  
DB 1 MFLPLGLGPPAAC 16  
1 ||||| |||||  
RESULT 10  
AAB61465  
ID AAB61465 standard; peptide; 14 AA.  
XX  
XX AAB61465;  
AC  
XX  
XX 04-APR-2001 (first entry)  
DT  
XX  
XX MANGO 245 signal peptide #1.  
DE  
XX  
XX TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
XX autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
XX pancreatic; skeletal; muscle.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200100672-A1.  
PN  
XX  
XX 04-JAN-2001.  
PD  
XX  
XX 29-JUN-2000; 2000MO-US18184.  
PF  
XX  
XX 29-JUN-1999; 99US-0342687.  
PR  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
PA  
XX  
XX Holtzman DA, Barnes TM, Fraser CC, Sharp JD;  
PI  
XX  
XX WPI, 2001-050127/06.  
DR  
XX  
XX Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
XX disorders (e.g. jaundice) -  
XX  
XX Disclosure; Page 242; 262pp; English.  
XX  
XX The present invention relates to cDNAs encoding TANGO 244,  
CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
CC The nucleic acids, proteins and protein modulators are useful for  
CC treating colonic disorders, inflammatory diseases, tumors,  
XX renal disorders, liver disorders, lung disorders, autoimmune diseases,

CC allergic diseases, cardiovascular diseases, brain disorders,  
CC degenerative diseases placental, pancreatic, skeletal and muscle  
XX disorders.  
XX  
XX Sequence 14 AA;  
SQ  
XX  
XX Query Match 82.8%; Score 72; DB 22; Length 14;  
Best Local Similarity 100.0%; Pred. No. 6.8e-05;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 MFLPLGLGPPAAC 14  
1 ||||| |||||  
DB 1 MFLPLGLGPPAAC 14  
1 ||||| |||||  
RESULT 11  
AAV52384  
ID AAV52384 standard; Protein; 115 AA.  
XX  
XX AAV52384;  
AC  
XX  
XX 09-FEB-2000 (first entry)  
DT  
XX  
XX Pseudomonas aeruginosa PilY2 protein.  
DE  
XX  
XX Pilus; motility; biofilm; biotic surface; abiotic surface;  
XX nutritional status; environment; gene expression; cellular physiology;  
XX antibiotic resistance; gene product; transcription;  
XX control; manufacturing; agriculture; healthcare; identification;  
XX screening; modulation; formation; colonisation; plant root;  
XX contact lens; medical implant; catheter wall; sewerage pipe;  
XX water pipe; lung infection.  
XX  
XX Pseudomonas aeruginosa.  
OS  
XX  
XX WO9955368-A1.  
PN  
XX  
XX 04-NOV-1999.  
PD  
XX  
XX 27-APR-1999; 99WO-US09034.  
PF  
XX  
XX 27-APR-1998; 98US-0083259.  
PR  
XX  
XX 02-OCT-1998; 98US-0102870.  
PR  
XX  
XX (HARD ) HARVARD COLLEGE.  
PA  
XX  
XX O'Toole GA, Kolter R;  
PI  
XX  
XX WPI, 2000-023264/02.  
DR  
XX  
XX Isolated surface attachment defective genes, useful for developing  
PT products which modulate bacterial biofilm formation -  
PT  
XX  
XX Claim 32; Page 149; 151pp; English.  
PS  
XX  
XX Sequences AAV52377-V52385 are involved in pilus synthesis and motility  
CC and are required for biofilm formation by Pseudomonas aeruginosa. The  
CC formation of bacterial biofilms occurs when bacteria make the transition  
CC from a planktonic (free swimming) existence to a lifestyle in which the  
CC microorganisms are firmly attached to biotic or abiotic surfaces. This  
CC transition is thought to be regulated in part by the nutritional status  
CC of the environment and is accompanied by a change in gene expression  
CC patterns, cellular physiology and higher resistance to antibiotics  
CC relative to planktonic bacterial counterparts. Pilus synthetic and  
CC motility genes, gene products, and transcriptional control regions may  
CC all be used for the control of biofilm formation in fields such as  
CC manufacturing, agriculture, and healthcare. The products may also be  
CC used to identify compounds which modulate biofilm formation. Compounds  
CC that stimulate biofilm formation can be used to improve colonisation of  
CC plant roots by beneficial bacteria (e.g., nitrogen fixing bacteria).  
CC Conversely, compounds that inhibit biofilm formation can be used to  
CC restrict the growth of bacteria on contact lenses, medical implants  
CC (e.g., artificial hips), the walls of catheters, fluid contacting

CC surfaces of industrial devices, water and sewerage pipes, and within the  
CC lungs of infected patients. The inhibitor compounds can also be applied  
CC to soil, seeds or plant roots to inhibit bacterial growth.

XX Sequence 115 AA;

Query Match 59.8%; Score 52; DB 21; Length 115;

Best Local Similarity 56.2%; Pred. No. 1.1; Mismatches 5; Indels 0; Gaps 0;

Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 MFLPLGLGPAACWA 16  
DB 3 VLPMLALAVPGCWA 18

RESULT 12  
ID ABB65553 standard; Protein; 1061 AA.

XX ABB65553;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 23451.

KM Drosophila; developmental biology; cell signalling; insecticide;  
KM Pharmaceutical.

XX Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR N-PSDB; ABL09656.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -

PS Disclosure; SEQ ID NO 23451; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB57737-AB872072).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1061 AA;

Query Match 56.3%; Score 49; DB 22; Length 1061;

Best Local Similarity 57.1%; Pred. No. 36;

Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 MFLPLGLGPAAC 14  
DB 740 LLPVLISLFGPSC 753

RESULT 13  
ID AAW61273 standard; Protein; 1180 AA.

XX AAW61273;

AC AAW61273;

DT 27-OCT-1998 (first entry)

DE Homo sapiens protocadherin-4.

XX pcch-4 gene; VE-cadherin; protocadherin-4; VE-cadherin-2;

KM vascular endothelial; angiogenesis; modulation; glaucoma; psoriasis;

KM inflammatory disease; organ transplantation; treatment; inhibition;

KM tumour; metastasis; rheumatoid arthritis; diagnosis; detection;

KM cell adhesion; atherosclerosis; myocardial ischaemia.

OS Homo sapiens.

PN WO9825946-A1.

PD 18-JUN-1998.

PF 05-NOV-1997; 97WO-US20006.

PR 12-DEC-1996; 96SE-0004731.

PA (IMCL-) IMCLONE SYSTEMS INC.

PI Dejana E, Telo P;

DR WPI; 1998-348441/30.

DR N-PSDB; AAV27593.

XX New isolated vascular endothelial cadherin-2 - used to develop  
PT products for modulating angiogenesis, e.g. for treating tumours,  
PT glaucoma, psoriasis, inflammatory diseases or organ transplantation

XX Claim 1; Page 31-37; 54pp; English.

XX The sequence is that of protocadherin-4 (pcch-4), or as it may  
CC alternatively be known, VE-cadherin-2. As a vascular endothelial  
CC cadherin it promotes cell-to-cell homotypic adhesion and its  
CC expression is upregulated in proliferating endothelial cells in  
CC comparison to resting cells. It can be used for inhibiting  
CC angiogenesis and inhibiting pathological conditions such as tumours,  
CC neovascular glaucoma, proliferative retinopathy including proliferative  
CC diabetic retinopathy, macular degeneration, hemangiomas, angiofibromas,  
CC and psoriasis. It may also be used for the prevention or inhibition of  
CC leukocyte infiltration, tumour cell metastasis, or endothelial  
CC permeability, as a vaccine and for making endothelial junctions more  
CC permeable to antigens, thus indicating use of the modifiers for  
CC treatment or prevention of acute and chronic inflammatory diseases,  
CC organ transplantation, myocardial ischaemia, atherosclerosis, rheumatoid  
CC arthritis and intestinal infection. It may also be used for detection,  
CC diagnosis and drug screening.

XX Sequence 1180 AA;

Query Match 56.3%; Score 49; DB 19; Length 1180;

Best Local Similarity 69.2%; Pred. No. 41;

Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 MFLPLGLGPA 13  
DB 4 LLPPLGLGPGS 16

RESULT 14  
ID AAUS2873 standard; Protein; 912 AA.

AC AAUS2873;  
XX  
DT 27-FEB-2002 (first entry)  
XX  
DE Propionibacterium acnes immunogenic protein #13769.  
XX  
KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;  
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
XX dermatological; osteopathic; neuroprotectant.  
XX  
OS Propionibacterium acnes.  
XX  
PN WO200181581-A2.  
XX  
PD 01-NOV-2001.  
XX  
PF 20-APR-2001; 2001WO-US12865.  
XX  
PR 21-APR-2000; 2000US-199047P.  
XX 02-JUN-2000; 2000US-208841P.  
PR 07-JUL-2000; 2000US-216747P.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhactia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX  
XX MPI; 2001-616774/71.  
DR N-PSDB; AAS59557.  
XX  
XX Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -  
XX  
XX Example 1; SEQ ID No 14068; 1069pp; English.  
XX  
XX Sequences AAUS3105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 912 AA;  
XX  
XX Query Match 55.2%; Score 48; DB 22; Length 912;  
XX Best Local Similarity 69.2%; Pred. No. 45;  
XX Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 4 LILGLGPAACWA 16  
|||||  
Db 373 LILGLGTPSSCWA 385

RESULT 15  
ABG17123  
ID ABG17123 standard; Protein; 167 AA.  
XX

AC ABG17123;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #17114.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSB-) HYSBQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX MPI; 2001-639362/73.  
DR N-PSDB; AAS81310.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX Claim 20; SEQ ID No 47482; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological actions  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 167 AA;  
XX  
XX Query Match 54.0%; Score 47; DB 22; Length 167;  
XX Best Local Similarity 58.8%; Pred. No. 10;  
XX Matches 10; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

OY 2 LPLLL-GILGPAACWA 16  
|||||  
Db 68 LPLLVLTGILFGWAECS 84

Search completed: January 12, 2004, 08:15:38  
Job time : 4.4118 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:15:44 ; Search time 1.9481 Seconds  
(without alignments)  
1655.023 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_1\_16  
Perfect score: 87  
Sequence: 1 MLPLILGLGPACWA 16

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747907 seqs, 201509753 residues  
Total number of hits satisfying chosen parameters: 747907

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	100.0	329	12 US-10-236-055A-14	Sequence 14, Appl
2	87	100.0	329	12 US-10-085-167-2	Sequence 2, Appl
3	73	89.9	326	12 US-10-236-055A-16	Sequence 16, Appl
4	49	56.3	1135	11 US-09-758-130B-42	Sequence 42, Appl
5	48	55.2	379	12 US-10-120-801-40	Sequence 8, Appl
6	47	54.0	656	10 US-09-554-000-8	Sequence 56, Appl
7	45	51.7	18	10 US-09-554-000-56	Sequence 41, Appl
8	45	51.7	347	12 US-10-120-801-41	Sequence 41, Appl
9	45	51.7	378	12 US-10-120-801-4	Sequence 4, Appl
10	45	51.7	416	12 US-10-316-253-2	Sequence 2, Appl
11	45	51.7	416	12 US-10-316-253-4	Sequence 4, Appl
12	45	51.7	416	12 US-10-316-253-6	Sequence 6, Appl
13	44	50.6	242	12 US-10-369-493-4780	Sequence 4780, Ap
14	44	50.6	242	12 US-10-369-493-7540	Sequence 7540, Ap
15	44	50.6	394	15 US-10-156-761-10413	Sequence 10413, A

16	44	50.6	615	10 US-09-908-193-26	Sequence 26, Appl
17	44	50.6	615	10 US-09-908-193-27	Sequence 27, Appl
18	44	50.6	1184	10 US-09-978-295A-425	Sequence 425, App
19	44	50.6	1184	10 US-09-978-697-425	Sequence 425, App
20	44	50.6	1184	10 US-09-978-192A-425	Sequence 425, App
21	44	50.6	1184	10 US-09-999-832A-425	Sequence 425, App
22	44	50.6	1184	11 US-09-978-189-425	Sequence 425, App
23	44	50.6	1184	11 US-09-978-608A-425	Sequence 425, App
24	44	50.6	1184	11 US-09-978-585A-425	Sequence 425, App
25	44	50.6	1184	11 US-09-978-191A-425	Sequence 425, App
26	44	50.6	1184	11 US-09-978-403A-425	Sequence 425, App
27	44	50.6	1184	11 US-09-978-564A-425	Sequence 425, App
28	44	50.6	1184	11 US-09-999-833A-425	Sequence 425, App
29	44	50.6	1184	11 US-09-981-915A-425	Sequence 425, App
30	44	50.6	1184	11 US-09-978-824-425	Sequence 425, App
31	44	50.6	1184	11 US-09-918-585A-425	Sequence 425, App
32	44	50.6	1184	11 US-09-978-423A-425	Sequence 425, App
33	44	50.6	1184	11 US-09-978-193A-425	Sequence 425, App
34	44	50.6	1184	11 US-09-999-830A-425	Sequence 425, App
35	44	50.6	1184	11 US-09-978-757A-425	Sequence 425, App
36	44	50.6	1184	11 US-09-978-187B-425	Sequence 425, App
37	44	50.6	1184	11 US-09-978-633A-425	Sequence 425, App
38	44	50.6	1184	12 US-09-978-375A-425	Sequence 425, App
39	44	50.6	1184	12 US-09-978-188A-425	Sequence 425, App
40	44	50.6	1184	12 US-09-978-298A-425	Sequence 425, App
41	44	50.6	1184	12 US-10-137-870-394	Sequence 394, App
42	44	50.6	1184	12 US-10-140-018-394	Sequence 394, App
43	44	50.6	1184	12 US-10-140-021-394	Sequence 394, App
44	44	50.6	1184	12 US-10-140-274-394	Sequence 394, App
45	44	50.6	1184	12 US-10-140-471-394	Sequence 394, App

#### ALIGNMENTS

RESULT 1  
US-10-236-055A-14  
Sequence 14, Application US/10236055A  
Publication No. US20030134328A1  
GENERAL INFORMATION:  
APPLICANT: Basham, Beth E.  
APPLICANT: Forsythe, Ian  
APPLICANT: Gorman, Daniel M.  
APPLICANT: Mattson, Jeanine  
APPLICANT: Mostrefi, Mehdiad  
APPLICANT: Patham, Chiriet  
TITLE OF INVENTION: MAMMALIAN GENES; RELATED REAGENTS  
FILE REFERENCE: DX01343K  
CURRENT APPLICATION NUMBER: US/10/236,055A  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/317,988  
PRIOR FILING DATE: 2001-09-06  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 14  
LENGTH: 329  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-236-055A-14  
Query Match 100.0%; Score 87; DB 12; Length 329;  
Best Local Similarity 100.0%; Pred. NO. 0.00012;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MLPLILGLGPACWA 16  
DB 1 MLPLILGLGPACWA 16  
RESULT 2  
US-10-085-167-2  
Sequence 2, Application US/10085167  
Publication No. US20030170781A1

```

; GENERAL INFORMATION:
; APPLICANT: Holloway, James L.
; APPLICANT: Lok, Si
; TITLE OF INVENTION: SECRETED PROTEIN ZACRPA
; FILE REFERENCE: 99-29
; CURRENT APPLICATION NUMBER: US/10/085,167
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: 60/141,928
; PRIOR FILING DATE: 1999-07-01
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 329
; TYPE: PRF
; ORGANISM: Homo sapiens
; US-10-085-167-2

Query Match      100.0%; Score 87; DB 12; Length 329;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 MFLPLGLGPAACWA 16
Db      1 MFLPLGLGPAACWA 16

RESULT 3
US-10-236-055A-16
; Sequence 16, Application US/10236055A
; Publication No. US20030134328A1
; GENERAL INFORMATION:
; APPLICANT: Basham, Beth E.
; APPLICANT: Forsythe, Ian
; APPLICANT: Gorman, Daniel M.
; APPLICANT: Matson, Jeanine
; APPLICANT: Moshrefi, Mehdiad
; APPLICANT: Parham, Christl
; TITLE OF INVENTION: MAMMALIAN GENES; RELATED REAGENTS
; FILE REFERENCE: DX01343K
; CURRENT APPLICATION NUMBER: US/10/236,055A
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/317,988
; PRIOR FILING DATE: 2001-09-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 326
; TYPE: PRF
; ORGANISM: Mus musculus
; US-10-236-055A-16

Query Match      83.9%; Score 73; DB 12; Length 326;
Best Local Similarity 87.5%; Pred. No. 0.011;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 MFLPLGLGPAACWA 16
Db      1 MFLPLGLGPAACWA 16

RESULT 4
US-09-759-130B-42
; Sequence 42, Application US/09759130B
; Publication No. US20030022279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarthy, Sean A
; APPLICANT: Fraser, Christopher C
; APPLICANT: Sharp, John D
; APPLICANT: Barnes, Thomas S
; APPLICANT: Kiser, Susan J
; APPLICANT: Mackay, Charles R
; APPLICANT: Myers, Paul S

; APPLICANT: Leiby, Kevin R
; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goodenall, Andrew
; APPLICANT: Holtzman, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; FILE REFERENCE: MPI00-5350NMIM
; CURRENT APPLICATION NUMBER: US/09/759,130B
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 1135
; TYPE: PRF
; ORGANISM: Mus sp.
; US-09-759-130B-42

Query Match      56.3%; Score 49; DB 11; Length 1135;
Best Local Similarity 69.2%; Pred. No. 95;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy      1 MFLPLGLGPAACWA 13
Db      4 LPLPLGLGPGS 16

RESULT 5
US-10-120-801-40
; Sequence 40, Application US/10120801
; Publication No. US20030203843A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shimkets, Richard
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Mehraban, Foad
; APPLICANT: Topper, James N.
; APPLICANT: Malvankar, Uriel
; APPLICANT: Wasserman, Scott
; APPLICANT: Edinger, Shlomit
; APPLICANT: Smithson, Glennda
; APPLICANT: Gunther, Erik
; APPLICANT: Komives, Laszlo
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-340
; CURRENT APPLICATION NUMBER: US/10/120,801
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 60/285748
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 60/286068
```



```
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/286292
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/288334
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/291241
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/322284
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/285609
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 40
; LENGTH: 379
; TYPE: PRT
; ORGANISM: mouse
US-10-120-801-40
```

```
Query Match      55.2%; Score 48; DB 12; Length 379;
Best Local Similarity 69.2%; Pred. No. 47;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy      1 MLPLGLGPPA 13
      :|||||:
Db      31 LPLPLGLMGPPA 43
```

```
RESULT 6
US-09-554-000-8
; Sequence 8, Application US/09554000
; Patent No. US20020165364A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: DETECTION OF ANALYTES
; FILE REFERENCE: 07257/042001
; CURRENT APPLICATION NUMBER: US/09/554,000
; CURRENT FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 08/818,252
; PRIOR FILING DATE: 1997-03-14
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 656
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-554-000-8
```

```
Query Match      54.0%; Score 47; DB 10; Length 656;
Best Local Similarity 83.3%; Pred. No. 1,le+02;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
Oy      2 LPLGLGLGPPA 13
      :|||||:
Db      5 VPLGLGLGAAA 16
```

```
RESULT 7
US-09-554-000-56
; Sequence 56, Application US/09554000
; Patent No. US20020165364A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: DETECTION OF ANALYTES
; FILE REFERENCE: 07257/042001
; CURRENT APPLICATION NUMBER: US/09/554,000
; CURRENT FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 08/818,252
; PRIOR FILING DATE: 1997-03-14
```

```
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 56
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-554-000-56
```

```
Query Match      51.7%; Score 45; DB 10; Length 18;
Best Local Similarity 83.3%; Pred. No. 7;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
Oy      2 LPLGLGLGPPA 13
      :|||||:
Db      5 VPLGLGLGAAA 16
```

```
RESULT 8
US-10-120-801-41
; Sequence 41, Application US/10120801
; Publication No. US20030203843A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shimkets, Richard
; APPLICANT: Padigaru, Murallidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Menzaban, Fuad
; APPLICANT: Topper, James N.
; APPLICANT: Malyankar, Uriel
; APPLICANT: Wasserman, Scott
; APPLICANT: Edinger, Shlomit
; APPLICANT: Smithson, Glenda
; APPLICANT: Gunther, Erik
; APPLICANT: Komuves, Laszlo
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-340
; CURRENT APPLICATION NUMBER: US/10/120,801
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 60/285748
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 60/286068
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/286292
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/288334
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/291241
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/322284
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/285609
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41
; LENGTH: 347
; TYPE: PRT
; ORGANISM: human
US-10-120-801-41
```

```
Query Match      51.7%; Score 45; DB 12; Length 347;
Best Local Similarity 69.2%; Pred. No. 1,le+02;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
Oy      1 MLPLGLGPPA 13
      :|||||:
Db      29 LPLPLGLMGPPA 41
```

```
RESULT 9
US-10-120-801-4
```

```
Sequence 4, Application US/10120801
Publication No. US2003020383A1
GENERAL INFORMATION:
APPLICANT: Pena, Carol
APPLICANT: Guo, Xiaojia
APPLICANT: Shinkens, Richard
APPLICANT: Padigaru, Muralidhara
APPLICANT: Kekuda, Ramesh
APPLICANT: Spytek, Kimberly
APPLICANT: Mehrahan, Fuad
APPLICANT: Topper, James N.
APPLICANT: Malvanekar, Uriel
APPLICANT: Wasserman, Scott
APPLICANT: Edinger, Shalomt
APPLICANT: Smithson, Glenda
APPLICANT: Gunther, Erik
APPLICANT: Komuves, Laszlo
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-340
CURRENT APPLICATION NUMBER: US/10/120,801
CURRENT FILING DATE: 2002-04-11
PRIOR APPLICATION NUMBER: 60/285748
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: 60/286068
PRIOR FILING DATE: 2001-04-24
PRIOR APPLICATION NUMBER: 60/286292
PRIOR FILING DATE: 2001-04-25
PRIOR APPLICATION NUMBER: 60/288334
PRIOR FILING DATE: 2001-05-03
PRIOR APPLICATION NUMBER: 60/291241
PRIOR FILING DATE: 2001-05-16
PRIOR APPLICATION NUMBER: 60/322284
PRIOR FILING DATE: 2001-09-14
PRIOR APPLICATION NUMBER: 60/285609
PRIOR FILING DATE: 2001-04-20
NUMBER OF SEQ ID NOS: 155
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 378
TYPE: PRT
ORGANISM: Homo sapiens
US-10-120-801-4

Query Match      51.7%  Score 45;  DB 12;  Length 378;
Best Local Similarity 69.2%;  Pred. No. 1.2e+02;
Matches 9;  Conservative 1;  Mismatches 3;  Indels 0;  Gaps 0;
```

```
QY      1  MLPLILGLGPAA 13
Db      29  LPLPLWLGLAGPCA 41
```

```
RESULT 10
US-10-316-253-2
Sequence 2, Application US/10316253
Publication No. US20030162706A1
GENERAL INFORMATION:
APPLICANT: The Procter & Gamble Company
APPLICANT: Peters, Kevin
APPLICANT: Thompson, Larry
APPLICANT: Wang, Feng
APPLICANT: Greis, Kenneth
TITLE OF INVENTION: Angiogenesis Modulating Proteins
FILE REFERENCE: 8865M
CURRENT APPLICATION NUMBER: US/10/316,253
CURRENT FILING DATE: 2002-12-10
PRIOR APPLICATION NUMBER: US 60/355,295
PRIOR FILING DATE: 2002-02-08
NUMBER OF SEQ ID NOS: 308
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 416
TYPE: PRT
```

```
ORGANISM: Rattus norvegicus
US-10-316-253-2
```

```
Query Match      51.7%  Score 45;  DB 12;  Length 416;
Best Local Similarity 83.3%;  Pred. No. 1.4e+02;
Matches 10;  Conservative 1;  Mismatches 1;  Indels 0;  Gaps 0;
```

```
QY      2  LPLILGLGPAA 13
Db      5  VPLILGLGLAA 16
```

```
RESULT 11
US-10-316-253-4
Sequence 4, Application US/10316253
Publication No. US20030162706A1
GENERAL INFORMATION:
APPLICANT: The Procter & Gamble Company
APPLICANT: Peters, Kevin
APPLICANT: Thompson, Larry
APPLICANT: Wang, Feng
APPLICANT: Greis, Kenneth
TITLE OF INVENTION: Angiogenesis Modulating Proteins
FILE REFERENCE: 8865M
CURRENT APPLICATION NUMBER: US/10/316,253
CURRENT FILING DATE: 2002-12-10
PRIOR APPLICATION NUMBER: US 60/355,295
PRIOR FILING DATE: 2002-02-08
NUMBER OF SEQ ID NOS: 308
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 416
TYPE: PRT
ORGANISM: Rattus norvegicus
US-10-316-253-4
```

```
Query Match      51.7%  Score 45;  DB 12;  Length 416;
Best Local Similarity 83.3%;  Pred. No. 1.4e+02;
Matches 10;  Conservative 1;  Mismatches 1;  Indels 0;  Gaps 0;
```

```
QY      2  LPLILGLGPAA 13
Db      5  VPLILGLGLAA 16
```

```
RESULT 12
US-10-316-253-6
Sequence 6, Application US/10316253
Publication No. US20030162706A1
GENERAL INFORMATION:
APPLICANT: The Procter & Gamble Company
APPLICANT: Peters, Kevin
APPLICANT: Thompson, Larry
APPLICANT: Wang, Feng
APPLICANT: Greis, Kenneth
TITLE OF INVENTION: Angiogenesis Modulating Proteins
FILE REFERENCE: 8865M
CURRENT APPLICATION NUMBER: US/10/316,253
CURRENT FILING DATE: 2002-12-10
PRIOR APPLICATION NUMBER: US 60/355,295
PRIOR FILING DATE: 2002-02-08
NUMBER OF SEQ ID NOS: 308
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 416
TYPE: PRT
ORGANISM: Rattus norvegicus
US-10-316-253-6
```

```
Query Match      51.7%  Score 45;  DB 12;  Length 416;
Best Local Similarity 83.3%;  Pred. No. 1.4e+02;
Matches 10;  Conservative 1;  Mismatches 1;  Indels 0;  Gaps 0;
```

OY 2 LPLGLGPA 13  
:|||||||  
DB 5 VPLGLGLA 16

## RESULT 13

US-10-369-493-4780  
; Sequence 4780, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 4780  
; LENGTH: 242  
; TYPE: PRT  
; ORGANISM: Burkholderia fungorum  
US-10-369-493-4780

Query Match 50.6%; Score 44; DB 12; Length 242;  
Best Local Similarity 64.3%; Pred. No. 1.1e+02;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 3 PLLGLGPA 16  
:|||||||  
DB 57 PLVVDLDPACSA 70

## RESULT 14

US-10-369-493-7540  
; Sequence 7540, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 7540  
; LENGTH: 242  
; TYPE: PRT  
; ORGANISM: Burkholderia cepacia  
US-10-369-493-7540

Query Match 50.6%; Score 44; DB 12; Length 242;  
Best Local Similarity 64.3%; Pred. No. 1.1e+02;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 3 PLLGLGPA 16  
:|||||||  
DB 57 PLVVDLDPACSA 70

RESULT 15  
US-10-156-761-10413

; Sequence 10413, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 10413  
; LENGTH: 394  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-10413

Query Match 50.6%; Score 44; DB 15; Length 394;  
Best Local Similarity 55.0%; Pred. No. 1.8e+02;  
Matches 11; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

OY 2 LPLGLGPA 15  
:|||||||  
DB 113 LPLGLGLTIVGIGVPA 132

Search completed: January 12, 2004, 08:33:40  
Job time: 4.9481 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:14 ; Search time 0.814371 Seconds  
(without alignments)  
831.284 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_1\_16  
Perfect score: 87  
Sequence: 1 MFLPILGLGPACMA 16

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PTCUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52.8	135	4	US-09-252-991A-21441	Sequence 21441, A
2	47.5	656	2	US-08-818-253-8	Sequence 8, Appli
3	47.5	656	3	US-08-818-253-8	Sequence 8, Appli
4	45.7	18	2	US-08-818-253-56	Sequence 56, Appli
5	45.7	18	2	US-08-818-253-61	Sequence 61, Appli
6	45.7	18	2	US-08-818-253-56	Sequence 56, Appli
7	44.4	318	1	US-08-375-962B-14	Sequence 14, Appli
8	44.4	318	1	US-08-940-424-3	Sequence 3, Appli
9	43.4	96	3	US-08-465-343A-11	Sequence 11, Appli
10	42.4	109	4	US-09-199-637A-389	Sequence 389, App
11	41.1	90	4	US-09-252-991A-19814	Sequence 19814, A
12	41.1	277	2	US-08-147-784-2	Sequence 2, Appli
13	41.1	277	3	US-08-195-967-2	Sequence 2, Appli
14	41.1	277	3	US-09-006-353A-12	Sequence 12, Appli
15	41.1	277	3	US-08-472-940-2	Sequence 2, Appli
16	41.1	277	4	US-09-573-986-12	Sequence 12, Appli
17	41.1	277	4	US-09-880-933-2	Sequence 2, Appli
18	41.1	277	4	US-09-804-200-2	Sequence 2, Appli
19	41.1	355	4	US-09-325-932A-191	Sequence 191, App
20	41.1	410	4	US-09-252-991A-11174	Sequence 11174, A
21	41.1	540	4	US-09-328-352-4514	Sequence 4514, Ap
22	41.1	1182	4	US-09-293-505-7	Sequence 7, Appli
23	41.1	1203	4	US-09-207-857-2	Sequence 2, Appli
24	41.1	1203	4	US-09-293-505-2	Sequence 2, Appli
25	41.1	1311	2	US-08-540-406-4	Sequence 4, Appli
26	41.1	1311	3	US-08-656-055-4	Sequence 4, Appli
27	41.1	1311	3	US-08-954-666-4	Sequence 4, Appli

28	41	47.1	1311	4	US-08-918-658-4	Sequence 4, Appli
29	41	47.1	1311	4	US-09-724-631-4	Sequence 4, Appli
30	41	47.1	1311	5	PCT-US95-13233-4	Sequence 4, Appli
31	40.5	46.6	176	4	US-09-252-991A-26925	Sequence 26925, A
32	40	46.0	218	4	US-09-252-991A-25415	Sequence 25415, A
33	40	46.0	584	2	US-08-845-295A-2	Sequence 2, Appli
34	40	46.0	584	3	US-09-140-933-2	Sequence 2, Appli
35	40	46.0	584	3	US-09-146-661-2	Sequence 2, Appli
36	40	46.0	584	3	US-09-150-515-2	Sequence 2, Appli
37	40	46.0	986	3	US-08-872-757-4	Sequence 4, Appli
38	40	46.0	986	3	US-08-872-757-4	Sequence 4, Appli
39	40	46.0	999	1	US-08-252-626A-2	Sequence 2, Appli
40	39.5	45.4	495	4	US-09-252-991A-19330	Sequence 19330, A
41	39	44.8	182	4	US-09-114-001C-5022	Sequence 5022, Ap
42	39	44.8	192	3	US-08-486-099-107	Sequence 107, App
43	39	44.8	192	3	US-08-360-107A-117	Sequence 117, App
44	39	44.8	192	3	US-08-484-223B-107	Sequence 107, App
45	39	44.8	192	3	US-08-919-597-107	Sequence 107, App

## ALIGNMENTS

RESULT 1  
US-09-252-991A-21441  
Sequence 21441, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 21441  
LENGTH: 135  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21441

Query Match 59.8%, Score 52, DB 4, Length 135,  
Best Local Similarity 56.2%, Pred. No. 0.58,  
Matches 9, Conservative 2, Mismatches 5, Indels 0, Gaps 0;

QY 1 MFLPILGLGPACMA 16  
DB 23 VLPMLALAVPLCMA 38

RESULT 2  
US-08-818-253-8  
Sequence 8, Application US/08818253  
Patent No. 5998204  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger Y.  
APPLICANT: Miyawaki, Atsushi  
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
NUMBER OF SEQUENCES: 61  
DETECTION OF ANALYTES  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FASTSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/818,253  
FILING DATE: 14-MAR-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Ph.D., Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07257/043001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 656 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-818-253-8

Query Match 54.0%; Score 47; DB 2; Length 656;  
Best Local Similarity 83.3%; Pred. No. 19;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LPLLLGLGPA 13  
DB 5 VPLLGLGLGAA 16

RESULT 3  
US-08-818-252-8  
Sequence 8, Application US/08818252B  
Patent No. 6197928  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger Y.  
APPLICANT: Miyawaki, Atsushi  
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
FILE REFERENCE: 07257/042001  
CURRENT APPLICATION NUMBER: US/08/818,252B  
CURRENT FILING DATE: 1997-03-14  
NUMBER OF SEQ ID NOS: 56  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 8  
LENGTH: 656  
TYPE: PRT  
ORGANISM: Aequorea victoria  
US-08-818-252-8

Query Match 54.0%; Score 47; DB 3; Length 656;  
Best Local Similarity 83.3%; Pred. No. 19;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LPLLLGLGPA 13  
DB 5 VPLLGLGLGAA 16

RESULT 4  
US-08-818-253-56  
Sequence 56, Application US/08818253  
Patent No. 5998204  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger Y.  
APPLICANT: Miyawaki, Atsushi  
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
FILE REFERENCE: 07257/042001  
CURRENT APPLICATION NUMBER: US/08/818,253  
CURRENT FILING DATE: 1997-03-14  
NUMBER OF SEQ ID NOS: 61  
SOFTWARE: FASTSEQ for Windows Version 2.0b  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 656 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-818-253-8

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FASTSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/818,253  
FILING DATE: 14-MAR-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Ph.D., Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07257/043001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 56:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-818-253-56

Query Match 51.7%; Score 45; DB 2; Length 18;  
Best Local Similarity 83.3%; Pred. No. 0.8;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LPLLLGLGPA 13  
DB 5 VPLLGLGLGAA 16

RESULT 5  
US-08-818-253-61  
Sequence 61, Application US/08818253  
Patent No. 5998204  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger Y.  
APPLICANT: Miyawaki, Atsushi  
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR  
FILE REFERENCE: 07257/042001  
CURRENT APPLICATION NUMBER: US/08/818,253  
CURRENT FILING DATE: 1997-03-14  
NUMBER OF SEQ ID NOS: 61  
SOFTWARE: FASTSEQ for Windows Version 2.0b  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 56:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-818-253-56

```
REFERENCE/DOCKET NUMBER: 07257/043001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 18 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-818-253-61

Query Match      51.7%: Score 45; DB 2; Length 18;
Best Local Similarity 83.3%; Pred. No. 0.8;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 LPLLGLGLGPA 13
       :|||||
       5 VPLLGLGLGLAA 16

Db

RESULT 6
US-08-818-252-56
Sequence 56, Application US/08818252B
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
APPLICANT: Miyawaki, Atsushi
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
TITLE OF INVENTION: DETECTION OF ANALYTES
FILE REFERENCE: 07257/042001
CURRENT FILING DATE: 1997-03-14
CURRENT FILING DATE: 1997-03-14
NUMBER OF SEQ ID NOS: 56
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 56
LENGTH: 18
TYPE: PRT
ORGANISM: Rattus norvegicus
US-08-818-252-56

Query Match      51.7%: Score 45; DB 3; Length 18;
Best Local Similarity 83.3%; Pred. No. 0.8;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 LPLLGLGLGPA 13
       :|||||
       5 VPLLGLGLGLAA 16

Db

RESULT 7
US-08-375-962B-14
Sequence 14, Application US/08375962B
GENERAL INFORMATION:
APPLICANT: SIMON, ANDRAS, HELLMAN, ULF, WERNSTEDT,
APPLICANT: CHRISTER, ERIKSSON, ULF.
TITLE OF INVENTION: Isolated Nucleic Acid Molecule
TITLE OF INVENTION: Which Codes for A 32 kDa Protein Having 11-Cis Retinol
TITLE OF INVENTION: Dehydrogenase Activity, and Which Associates With P63,
TITLE OF INVENTION: a Portion of a Retinol Binding Protein Receptor
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect (ASCII standard)
```

```
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/375,962B
FILING DATE: 20-January-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/258,418
FILING DATE: 6-October-1994
ATTORNEY/AGENT INFORMATION:
NAME: Pasqualini, Patricia A.
REGISTRATION NUMBER: 34,894
REFERENCE/DOCKET NUMBER: LOD 5372.1 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 318 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: amino acid
FEATURE:
NAME/KEY: p32,11-cis retinol dehydrogenase
US-08-375-962B-14

Query Match      50.6%: Score 44; DB 1; Length 318;
Best Local Similarity 71.4%; Pred. No. 25;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      2 LPLLGLGLGPAACW 15
       :|||||
       3 LPLLGLVLMALM 16

Db

RESULT 8
US-08-940-424-3
Sequence 3, Application US/08940424A
GENERAL INFORMATION:
APPLICANT: Blauer, William S.
APPLICANT: Zolt, Roseann P.
APPLICANT: Gamble, Mary V.
APPLICANT: Metz, James R.
TITLE OF INVENTION: POTENT INHIBITORS OF HUMAN 9-cis RETINOL DEHYDROGENASE
FILE REFERENCE: 0575/54544
CURRENT FILING DATE: 1997-09-29
CURRENT FILING DATE: 1997-09-29
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 318
TYPE: PRT
ORGANISM: rat
US-08-940-424-3

Query Match      50.6%: Score 44; DB 3; Length 318;
Best Local Similarity 71.4%; Pred. No. 25;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      2 LPLLGLGLGPAACW 15
       :|||||
       3 LPLLGLVLMALM 16

Db

RESULT 9
US-08-465-343A-11
Sequence 11, Application US/08465343A
GENERAL INFORMATION:
APPLICANT: WYLER, David J.
APPLICANT: Prakash, Sadhana
APPLICANT: Zhang, Xiaoping
TITLE OF INVENTION: PSF-1 AND THE EARLY DETECTION
TITLE OF INVENTION: OF FIBROSIS
```

```

; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,343A
; FILING DATE: 05-JUN-1995
; APPLICATION DATA:
; APPLICATION NUMBER: 08/395,674
; FILING DATE: 28-FEB-1995
; APPLICATION NUMBER: 08/152,904
; FILING DATE: 15-NOV-1993
; APPLICATION NUMBER: 07/840,426
; FILING DATE: 24-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 00398/096002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 96 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-465-343A-11

Query Match          49.4%; Score 43; DB 3; Length 96;
Best Local Similarity 90.0%; Pred. No. 9.9;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MPELLIGLIG 10
Db 1 MPELLIGLYG 10

RESULT 10
US-09-199-637A-389
; Sequence 389, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 389
; LENGTH: 109

```

```

; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-199-637A-389

Query Match          48.3%; Score 42; DB 4; Length 109;
Best Local Similarity 60.0%; Pred. No. 16;
Matches 9; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

Qy 7 GILGPA-----CM 15
Db 71 GILGPAETPLSLCW 85

RESULT 11
US-09-252-991A-19814
; Sequence 19814, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT FILING DATE: 1999-02-18
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19814
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-19814

Query Match          47.1%; Score 41; DB 4; Length 90;
Best Local Similarity 69.2%; Pred. No. 19;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 LPLILGLIGPAAAC 14
Db 30 LPLMFCILGLAAC 42

RESULT 12
US-08-147-784-2
; Sequence 2, Application US/08147784
; Patent No. 5821332
; GENERAL INFORMATION:
; APPLICANT: Godfrey, Wayne
; APPLICANT: Buck, David
; APPLICANT: Engleman, Edgar G.
; TITLE OF INVENTION: RECEPTOR ON THE SURFACE OF ACTIVATED
; TITLE OF INVENTION: CD4+ T-CELLS: ACT-4
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: 379 Lytton Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: US
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/147,784
; FILING DATE: 03-NOV-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, William M

```



REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 05490A-220  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 326-2422  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 277 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-147-784-2

Query Match 47.1%; Score 41; DB 2; Length 277;  
Best Local Similarity 75.0%; Pred. No. 63;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 LPLVLGLGPPA 13  
DB 221 LGLVGLGLGPPA 232

RESULT 13  
US-08-195-967-2  
Sequence 2, Application US/08195967  
Patent No. 6242566  
GENERAL INFORMATION:  
APPLICANT: Godfrey, Wayne  
APPLICANT: Engleman, Edgar G.  
TITLE OF INVENTION: LIGAND (ACT-4-L) TO A RECEPTOR ON THE SURFACE OF ACTIVATED  
TITLE OF INVENTION: CD4+ T-CELLS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESS: Townsend and Townsend Khourie and Crew  
STREET: 379 Lytton Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: US  
ZIP: 94301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/195,967  
FILING DATE: 10-FEB-1994  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William M  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 05490A-230  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 326-2422  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 277 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-195-967-2

Query Match 47.1%; Score 41; DB 3; Length 277;  
Best Local Similarity 75.0%; Pred. No. 63;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 LPLVLGLGPPA 13  
DB 221 LGLVGLGLGPPA 232

RESULT 14

US-09-006-353A-12  
Sequence 12, Application US/09006353A  
Patent No. 6261801

GENERAL INFORMATION:  
APPLICANT: WEI, YING-FEI  
APPLICANT: YU, GUO-LIANG  
APPLICANT: GENTZ, REINER  
APPLICANT: RUBEN, STEVEN  
TITLE OF INVENTION: TUMOR NECROSIS FACTOR RECEPTOR 5  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESS: HUMAN GENOME SCIENCES, INC.  
STREET: 9410 KEY WEST AVENUE  
CITY: ROCKVILLE  
STATE: MD  
COUNTRY: US  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/006,353A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BROOKES, ANDERS A  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: P341  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 277 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-006-353A-12

Query Match 47.1%; Score 41; DB 3; Length 277;  
Best Local Similarity 75.0%; Pred. No. 63;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 LPLVLGLGPPA 13  
DB 221 LGLVGLGLGPPA 232

RESULT 15  
US-08-472-940-2  
Sequence 2, Application US/08472940  
Patent No. 6277962  
GENERAL INFORMATION:  
APPLICANT: Godfrey, Wayne  
APPLICANT: Buck, David  
APPLICANT: Engleman, Edgar G.  
TITLE OF INVENTION: RECEPTOR ON THE SURFACE OF ACTIVATED  
TITLE OF INVENTION: CD4+ T-CELLS: ACT-4  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESS: Townsend and Townsend Khourie and Crew  
STREET: 379 Lytton Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: US  
ZIP: 94301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/472,940  
 FILING DATE: 06-JUN-1995  
 CLASSIFICATION: 530  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/147,784  
 FILING DATE: 03-NOV-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Smith, William M  
 REGISTRATION NUMBER: 30,223  
 REFERENCE/DOCKET NUMBER: 05490A-220  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 326-2400  
 TELEFAX: (415) 326-2422  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 277 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULAR TYPE: protein  
 US-08-472-940-2

Query Match 47.1%; Score 41; DB 3; Length 277;  
 Best Local Similarity 75.0%; Pred. No. 63;  
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 LPLLLGLGPPAA 13  
 Db 221 LGLVGLGLGPLA 232

Search completed: January 12, 2004, 08:20:07  
 Job time: 2.81437 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:09 ; Search time 8.27744 Seconds  
(without alignments)  
1661.397 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_17\_159  
Sequence: 1 LGPRGSGSELRSAPFSAAR.....PHVALGAPGATFGSLVYAD 143

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR.76:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	191	25.6	244	JC4708	gelatin-binding 28
2	189	25.4	219	T14782	hypothetical prote
3	168.5	22.6	253	C1HU0B	complement subcomp
4	163.5	21.9	253	S49158	complement protein
5	160	21.5	245	C1HU0C	complement subcomp
6	156	20.9	245	S19018	complement subcomp
7	152.5	20.5	246	S29338	complement subcomp
8	150	20.1	224	A60032	cerebellin-like g1
9	149.5	20.1	253	I49560	complement C1q B c
10	149	20.0	193	A37873	cerebellin precurs
11	148.5	19.9	674	S23287	collagen alpha 1(X
12	148	19.9	674	S13301	collagen alpha 1(X
13	147	19.7	245	C1HU0A	complement subcomp
14	145	19.5	680	CGHUID	collagen alpha 1(X
15	144.5	19.4	215	C48150	hypothetical relate
16	144	19.3	1228	A57384	multimerin, endoch
17	139.5	18.7	635	A57131	collagen alpha 2(V
18	139.5	18.7	744	A34246	collagen alpha 1(V
19	139	18.7	423	A55797	collagen precursor
20	137.5	18.5	744	S15435	collagen alpha 1(V
21	136.5	18.3	215	B48150	hypothetical relate
22	135	18.3	743	S23779	collagen alpha 1(V
23	135.5	18.2	170	B57131	collagen alpha 2(V
24	134.5	18.1	744	S23288	collagen alpha 1(V
25	132	17.7	680	S31216	collagen alpha 1(X
26	121.5	16.3	992	T08772	hypothetical prote
27	113.5	15.2	196	A48150	hypothetical relate
28	75	10.2	366	T52655	thiosulfate sulfur
29	76	10.2	483	E75384	conserved hypothet

30	75	10.1	272	2	B38453	anaerobic sulfite
31	74	9.9	272	2	AD0825	anaerobic sulfite
32	74	9.9	564	2	E87631	acyl-CoA dehydroge
33	73	9.8	413	2	S22578	translation initia
34	73	9.8	789	2	S33056	probable infected
35	71	9.5	748	2	T32654	hypothetical prote
36	70	9.4	314	2	T35241	hypothetical prote
37	69	9.3	303	2	B83336	hypothetical prote
38	69	9.3	355	2	S52022	translation initia
39	69	9.3	413	2	S55898	translation initia
40	69	9.3	413	2	S52018	translation initia
41	69	9.3	413	2	S52019	translation initia
42	69	9.3	413	2	S52023	translation initia
43	69	9.3	571	2	H83696	DNA mismatch repai
44	69	9.3	797	2	A87247	probable secreted
45	68.5	9.2	333	2	E84476	probable Athila re

#### ALIGNMENTS

##### RESULT 1

JC4708  
gelatin-binding 28k protein precursor - human

N/Alternate names: adipose specific collagen-like factor

C/Species: Homo sapiens (man)

C/Date: 10-May-1996 #sequence\_revision 19-Jul-1996 #text\_change 20-Sep-1999

C/Accession: JC4708; JC4944

R/Mada, K.; Okubo, K.; Shimomura, I.; Funahashi, T.; Matsuzawa, Y.; Matsubara, K.

Biochem. Biophys. Res. Commun. 221, 286-289, 1996

A/Title: cDNA cloning and expression of a novel adipose specific collagen-like factor,

A/Reference number: JC4708; PMID:96224171; PMID:8619847

A/Accession: JC4708

A/Molecule type: mRNA

A/Residues: 1-244 <MAE>

A/Cross-references: DDBJ:D45371; NID:G871886; PIDN:BAA08227.1; PID:G871887

A/Experimental source: adipose tissue

R/Makano, Y.; Tohe, T.; Choi-Makura, N.H.; Mazda, T.; Tomita, M.

J. Biochem. 120, 803-812, 1996

A/Title: Isolation and characterization of GBP28, a novel gelatin-binding protein purif

A/Reference number: JC4944; PMID:97103474; PMID:8947845

A/Accession: JC4944

A/Molecule type: protein

A/Residues: 19-38;93-100;101-112;135-149;173-178 <NAK>

C/Comment: This protein is an endogenous factor that binds with a collagen-like domain.

C/Genetics:

A/Gene: apm1

C/Superfamily: unassigned collagens; complement C1q carboxyl-terminal homology

C/Keywords: adipose tissue; glycoprotein; hydroxyproline

F/1-18/Domain: signal sequence #status predicted <SIG>

F/19-244/Product: gelatin-binding 28kDa protein #status experimental <MAT>

F/42-107/Region: collagen-like

F/114-241/Domain: complement C1q carboxyl-terminal homology <CIQ>

F/95/Modified site: 4-hydroxyproline (Pro) #status experimental

F/230/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 25.6% Score 191; DB 2; Length 244;

Best Local Similarity 31.8% Pred. NO. 6.3e-12; Indels 18; Gaps 6;

Matches 47; Conservative 29; Mismatches 54;

QY 5 PGSGSELRSAPFSAATTPLEGTSEMAVFDKVVYVIGDPVATQFCRVGAYFSF 64

Db 104 PGGAAYVVSASVGLTETVT-IPNPIRFTKIFVYQNHGSGSTCKFNCNPGLYFAY 162

QY 65 TAGKAPH-----KSLVLMVRNDEVQALAFDORRFGARRAASGAMQLDYGDTWLR 119

Db 163 -----HITVYMKDVSVSLFK-KDKRAMEFTYQYOYENNVDQ-ASGSLVLLHLEVGQVWQ 214

QY 120 LHGAHF-----VALGAPGATFGSLVYAD 143

Db 215 VVGEGRNGLYADNDNDSTFTGFLYHD 242

## RESULT 2

hypothetical protein DKFZ586B0621.1 - human (fragment)  
T14782  
C/Species: Homo sapiens (man)  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 02-Sep-2000  
C/Accession: T14782  
R/Octomwelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, August 1999  
A/Reference number: Z18184  
A/Accession: T14782  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-219 <OTT>  
A/Cross-references: EMBL:AL110261  
A/Experimental source: adult uterus; clone DKFZ586B0621  
C/Genetics:  
A/Note: DKFZ586B0621.1  
C/Superfamily: complement C1q carboxyl-terminal homology

Query Match 25.4%; Score 189; DB 2; Length 219;  
Best Local Similarity 35.3%; Pred. No. 8.9e-12;  
Matches 53; Conservative 28; Mismatches 57; Indels 12; Gaps 6;  
Qy 2 GPTGPGSSSL--SSAFSAART--TPLEGTSEMAVTPDKVYVNIIGSDPVAATGQRCRVPG 58  
Db 66 GPTGPGGCSVPSPSAFSAKSESRRVPPSDAPLPBFDVLVNEGHDAVATGKFTCPVG 125  
Qy 59 AYFFSFTAGKAPHKSLSVMLVRNDEVOL--AFDEQRRPARRAASOSAMLQIDYDPTV 116  
Db 126 VYFVAVHA--TVYRASLPQDLVKNGEISLAFQFGCGMKPA---SLSCGAVRLPEPDV 181  
Qy 117 WLRLHGAPH--VALGAGATFSGLYVAD 143  
Db 182 WVQGVGDYIGIYASIKTDSFSGFLVSD 211

## RESULT 3

CIH0B  
complement subcomponent C1q chain B precursor (validated) - human  
N/Alternate names: complement subcomponent C1q beta chain  
C/Species: Homo sapiens (man)  
C/Date: 22-May-1981 #sequence\_revision 31-May-1996 #text\_change 08-Dec-2000  
C/Accession: B03422; A23422; B90304; A90301; B90315; A03206  
R/Reid, K.B.M.  
Biochem. J. 231, 729-735, 1985  
A/Title: Molecular cloning and characterization of the complementary DNA and gene coding  
A/Reference number: A23422; MUID:86076906; PMID:3000358  
A/Accession: B23422  
A/Molecule type: DNA  
A/Residues: 'HS', 1-32 <REI>  
A/Note: the authors translated the codon AGT for the second position as Arg; they were u  
A/Accession: A23422  
A/Molecule type: mRNA  
A/Residues: 28-253 <REI>  
A/Cross-references: EMBL:X03084  
A/Note: the authors translated the codon ACA for residue 46 as Ile  
R/Reid, K.B.M.  
Biochem. J. 179, 367-371, 1979  
A/Title: Complete amino acid sequences of the three collagen-like regions present in sub  
A/Reference number: A90304; MUID:80020137; PMID:486087  
A/Accession: B90304  
A/Molecule type: protein  
A/Residues: 'E', 29-84, 'D', 86-99, 'P', 101-135 <RES>  
R/Reid, K.B.M.; Thompson, E.O.P.  
Biochem. J. 173, 863-868, 1978  
A/Title: Amino acid sequence of the N-terminal 108 amino acid residues of the B chain of  
A/Reference number: A90301; MUID:79041552; PMID:708376  
A/Accession: A90301  
A/Molecule type: protein  
A/Residues: 28-99, 'P', 101-195 <RE3>  
R/Reid, K.B.M.; Gagnon, J.; Frampton, J.  
Biochem. J. 203, 559-569, 1982  
A/Title: Completion of the amino acid sequences of the A and B chains of subcomponent C1

A/Reference number: A90315; MUID:82283890; PMID:6981411

A/Accession: B90315  
A/Molecule type: protein  
A/Residues: 136-253 <RE4>  
A/Note: 176-GLX may also be present  
C/Comment: The first component of complement is a calcium-dependent complex of the three  
activation of C1r (enzyme), C1s (proenzyme), and the other eight components of complement.  
C/Comment: The C1q subcomponent is composed of nine subunits, six of which are disulfide  
(see PIR:CIH0C) chain. Equimolar amounts of the A, B, and C chains are found after red  
C/Genetics:  
A/Gene: GDB:CIQB  
A/Cross-references: GDB:119043; OMIM:120570  
A/Map position: 1p36.3-1p34.1  
C/Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal ho  
C/Superfamily: complement pathway; glycoprotein; heterodimer; hydroxylysine; hydroxyprolin  
F:1-27/Domain: signal sequence #status predicted <SIG>  
F:28-253/Product: complement subcomponent C1q chain B #status experimental <MAT>  
F:33-116/Domain: collagenous, triple helix <COL>  
F:123-249/Domain: complement C1q carboxyl-terminal homology <CIQ>  
F:28/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experime  
F:33/38,41,53,56,65,83,86,101,104,107/Modified site: 4-hydroxyproline (Pro) #status exp  
F:59,62,77,92,98,110/Modified site: 5-hydroxylysine (Lys) #status experimental  
F:59,62,98,110/Binding site: carboxylate (Lys) (covalent) #status experimental

Query Match 22.6%; Score 168.5; DB 1; Length 253;  
Best Local Similarity 27.7%; Pred. No. 1.3e-09;  
Matches 44; Conservative 34; Mismatches 62; Indels 19; Gaps 4;  
Qy 1 LGPTPGGS-----SELSAFSAARTTPLEGTSEMAVTPDKVYVNIIGSDPD 46  
Db 95 MGKPGKGAGAPGAPKPKESGDYKATOKIAPSAATRTTNVPRROQTIRFDVITNNMNYE 154  
Qy 47 VANGQFRCRVGAYFFSFTAGKAPHKSLSVMLVRNDEVOLAFDEQRRPARRAASQSA 106  
Db 155 PRSGKFTCKVPGLYFTYHA--SSKGNLCVNLKRGERRAKVYTPCDYAINTFQVTTGCM 212  
Qy 107 MLOLDYGVTVLRLHGAPHYALGAPGAT--FSGLYVAD 143  
Db 213 VLKLGQENVFLQATD-KNSLKGEGANSJFSGFLFPD 250

## RESULT 4

S49158  
complement protein C1q beta chain precursor - rat  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 16-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 20-Aug-1999  
C/Accession: S49158  
R/Schwaebler, W.; Petry, F.; Loos, M.  
submitted to the EMBL Data Library, March 1993  
A/Description: cDNA cloning and expression of the mRNA encoding the B-chain of rat C1q.  
A/Reference number: S49158  
A/Accession: S49158  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-253 <SCH>  
A/Cross-references: EMBL:X71127; NID:9510191; PID:CAA50440.1; PID:9510192  
C/Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal ho  
F:121-249/Domain: complement C1q carboxyl-terminal homology <CIQ>  
Query Match 21.9%; Score 163.5; DB 2; Length 253;  
Best Local Similarity 29.2%; Pred. No. 4.2e-09;  
Matches 45; Conservative 36; Mismatches 58; Indels 15; Gaps 7;  
Qy 2 GPTGPG-----GSSELSA--AFSAARTTPLEGTSEMAVTPDKVYVNIIGSDPVAATGQF 52  
Db 100 GP-GGPRGPKGSGSDYKATOKIAPSAATRTTNVPRROQTIRFDVITNNMNYEPRSGKF 158  
Qy 53 RCRVGAFFSFTAGKAPHKSLSVMLVRNDEVOLAFDEQRRPARRAASOSAMLQID 111  
Db 159 TCKVPGLYFTYHA--SSKGNLCVNLKRGERRAKVYTPCDYAINTFQVTTGCVLKL 217  
Qy 112 YGDTVTLRLHGAPHYALGAPGAT--FSGLYVAD 143

Db 218 QEBVHLQATD-KNSLLGVEGANSIFTGFLFPD 250

## RESULT 5

C1HUC

Complement subcomponent C1q chain C precursor - human

N:Alternate names: complement subcomponent C1q gamma chain

C:Species: Homo sapiens (man)

C:Date: 22-May-1991 #sequence\_revision 31-May-1996 #text\_change 22-May-1998

C:Accession: S14351; A03207

R:Sellar, G.C.; Blake, D.J.; Reid, K.B.M.

Biochem. J. 274, 481-490, 1991

A:Title: Characterization and organization of the genes encoding the A-, B- and C-chains

A:Reference number: S14350; PMID:91174759; PMID:1706597

A:Accession: S14351

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-245 &lt;SEL&gt;

R:Reid, K.B.M.

Biochem. J. 179, 367-371, 1979

A:Title: Complete amino acid sequences of the three collagen-like regions present in sub

A:Reference number: A90304; PMID:80020137; PMID:486087

A:Accession: A03207

A:Molecule type: protein

A:Residues: 29-56, 'P', 58-65, 'K', 67-71, 'P', 73-83, 'K', 85-86, 'D', 88-89, 'N', 91-122 &lt;REI&gt;

C:Comment: The first component of complement is a calcium-dependent complex of the three

Ivation of C1r (enzyme), C1s (proenzyme), and the other eight components of complement.

C:Comment: The C1q subcomponent is composed of nine subunits, six of which are disulfide

dimers of the C chain. Equimolar amounts of the A, B, and C chains are found after red

C:Genetics:

A:Gene: GDB:C10G

A:Cross-references: GDB:128132; OMIM:120575

A:Map position: 1p36.3-1p34.1

A:Intons: 60/3

C:Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal hom

C:Keywords: complement pathway; glycoprotein; homodimer; hydroxyllysine; hydroxyproline;

F:1-28/Domain: signal sequence #status predicted &lt;SIG&gt;

F:29-245/Product: complement subcomponent C1q chain B #status predicted &lt;MNT&gt;

F:131-114/Domain: collagenous, triple helix &lt;COL&gt;

F:131-244/Domain: complement C1q carboxyl-terminal homology &lt;C1Q&gt;

F:33/Disulfide bonds: interchain #status experimental

F:36,39,42,45,54,63,81,93,96,99,105/Modified site: 4-hydroxyproline (Pro) #status experi

F:57,72,75/Modified site: 5-hydroxylysine (Lys) #status experimental

F:75/Binding site: carbohydrate (Lys) (covalent) #status experimental

Query Match

Best Local Similarity 21.5%; Score 160; DB 1; Length 245;

Matches 44; Conservative 28; Mismatches 63; Indels 18; Gaps 6;

2 GPTPGPG-----SSELRSAFSAARTTPLEGTSEMAVTFDKYVYVNIIGDFVATGQFR 53

Db 100 GPWGIPGPGEGRGYKQFQSVFTVTRQTHQPPANSLIRNAVLTNPQSGYDSTGKFT 159

QY 54 CRVGAFFFTTAKAPHKSLSVMLVNRNDEVQALAFDEQRPPARRASQAMQLDYG 113

Db 160 CKVPGLYFYVYHSHHT--ANLCVLLYRSQVYVFCGHTSK---TQVNSGVALRLQVG 214

QY 114 DTWMLRLHGAPHY-ALGAPG--ATPSGYLYVAD 143

Db 215 EEWMLAVN--DYDMVGIGSDSVFSGFLFPD 245

## RESULT 6

S19018

Complement subcomponent C1q chain A precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 20-Aug-1999

C:Accession: S19018

R:Petry, F.; Reid, K.B.M.; Loos, M.

J. Immunol. 147, 3988-3993, 1991

A:Title: Gene expression of the A- and B-chain of mouse C1q in different tissues and the

A:Accession: S19018

A:Molecule type: mRNA

A:Residues: 1-245 &lt;PEP&gt;

A:Cross-references: EMBL:X58861; NID:950226; PIDN:CAA41664.1; PID:950227

C:Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal ho

F:116-243/Domain: complement C1q carboxyl-terminal homology &lt;C1Q&gt;

Query Match

Best Local Similarity 20.9%; Score 156; DB 2; Length 245;

Matches 48; Conservative 15; Mismatches 58; Indels 32; Gaps 4;

5 PGSSSLRSASFSAARTTPLEGTSEMAVTFDKYVYVNIIGDFVATGQFRCPVAPYFSP 64

Db 106 PGNTRDQPRDAFSAIRQNPW--TLGNVYIFDKVLTNGESPQYQNHRTGRFICAVPGFYFNF 163

QY 65 -----TAGKAPHKSLSVMLVNRNDEVQALAFDEQRPPARRASQAMQLD 110

Db 164 QVTSKMDLCFISSSGGQFRDLSFSTNTNKKGLFQVLA-----GGTVLQL 209

QY 111 DYGDVWMLRLHGAPH--YALGAPGATFGYLYV 141

Db 210 RRGDEVWIERKDPKAGRTYGTEDSIFSGFLIF 242

QY 111 DYGDVWMLRLHGAPH--YALGAPGATFGYLYV 141

Db 210 RRGDEVWIERKDPKAGRTYGTEDSIFSGFLIF 242

## RESULT 7

S29328

Complement subcomponent C1q chain C - mouse

C:Species: Mus musculus (house mouse)

C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 20-Aug-1999

C:Accession: S29328

R:Petry, F.; Reid, K.B.M.; Loos, M.

Eur. J. Biochem. 209, 129-134, 1992

A:Title: Isolation, sequence analysis and characterization of cDNA clones coding for th

cerebellin.

A:Reference number: S29328; PMID:93011118; PMID:1396691

A:Accession: S29328

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-246 &lt;PEP&gt;

A:Cross-references: EMBL:X66295; NID:950228; PIDN:CAA46993.1; PID:950229

C:Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal ho

F:112-245/Domain: complement C1q carboxyl-terminal homology &lt;C1Q&gt;

Query Match

Best Local Similarity 20.5%; Score 152.5; DB 2; Length 246;

Matches 44; Conservative 29; Mismatches 59; Indels 21; Gaps 5;

2 GPTPGPG-----SSELRSAFSAARTTPLEGTSEMAVTFDKYVYVNIIGDFVATGQFR 56

Db 104 GPGEDEGVEGRYKQKQSVFTVTRQTHQPPANSLIRNAVLTNPQSGYDSTGKFTCEV 163

QY 57 PGAYFF-----SFTAGKAPHKSLSVMLVNRNDEVQALAFDEQRPPARRASQAMQLDYG 112

Db 164 PGLYFYFYVYTSHTANLCVHLINLTARVASFCD-----HMFNSKQVSSGALLRLQR 214

QY 113 GDTWMLRLHGAPHYALGAPG--ATPSGYLYVAD 143

Db 215 GDEVWLSVNDY-NGWVGIEBSNSVFSGLFPD 246

## RESULT 8

A60032

cerebellin-like glycoprotein - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 03-Mar-1993 #sequence\_revision 03-Mar-1993 #text\_change 31-Mar-2000

C:Accession: A60032

R:Wada, C.; Ohtani, H.

Brain Res. Mol. Brain Res. 9, 71-77, 1991

A:Title: Molecular cloning of rat cerebellin-like protein cDNA which encodes a novel me

A:Reference number: A60032; PMID:91203483; PMID:1850079

A:Accession: A60032

A:Status: not compared with conceptual translation

A:Molecule type: mRNA



A/Accession: S7711  
 A/Molecule type: protein  
 A/Residues: 104-112, 'X', 114-117, 453-466 <NIN2>  
 C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
 C/Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer; hydroxyproline;  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:547-673/Domain: complement C1q carboxyl-terminal homology <C1Q>  
 F:453-456/Modified site: hydroxyproline (Pro) #status experimental  
 F:611/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 19.9%; Score 148.5; DB 2; Length 674;  
 Best Local Similarity 27.6%; Pred. No. 4.3e-07;  
 Matches 47; Conservative 23; Mismatches 65; Indels 35; Gaps 5;

QY 2 GPTGPGSSSELRSAFSAARTPLTSTSEM-----AV 32  
 DB 504 GPPGPGGOSTIDEGYKSESELSGMSFPMKAGANQALGMPVSAFTVILSKAYPAATVPI 563

QY 33 TFDKVVYVIGGDFVATGFCRCVPGAYFFSFTAGKAPHSLSVWLVNRRDEVQALAFDE 92  
 DB 564 KFDKLVNRQOHYDPRITGFCRIPLGLYFSYHV-HAAGTWNVALYNGSPVM-XTYDE 621

QY 93 QRRGARRAASQASAMLDYDGTWVLRHGAAPH--VALGAPGATFSGYL 139  
 DB 622 YQK-GYLDQASGSAVIDLMENDQVWLQLPNSESNGLYSEYVHSSFGFL 670

RESULT 12  
 S13301  
 collagen alpha 1(X) chain precursor - bovine  
 C/Species: Bos primigenius taurus (cattle)  
 C/Date: 21-Nov-1993 #sequence\_revision 23-Feb-1996 #text\_change 13-Aug-1999  
 C/Accession: S13301  
 R/Thomas, J.T.; Kean, A.P.L.; Grant, M.E.; Boot-Handford, R.P.  
 Biochem. J. 273, 141-148, 1991  
 A/Title: Isolation of cDNAs encoding the complete sequence of bovine type X collagen. By  
 A/Reference number: S13301; MUID:91113131; PMID:1703407  
 A/Accession: S13301  
 A/Molecule type: mRNA  
 A/Residues: 1-674 <THO>  
 A/Cross-references: EMBL:X53556; NID:G263; PIDN:CAAJ7624.1; PID:G264  
 C/Genetics:  
 A/Gene: COL10A1  
 C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
 C/Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-674/Product: collagen alpha 1(X) chain #status predicted <MAT>  
 F:547-673/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 19.9%; Score 148; DB 2; Length 674;  
 Best Local Similarity 31.5%; Pred. No. 4.9e-07;  
 Matches 52; Conservative 19; Mismatches 64; Indels 30; Gaps 7;

QY 2 GPTGPGSSSELRSAFSAARTPLTSTSEMAYT-----FDKYYV 39  
 DB 511 GPPGPGGVALPBDVFKAGORPVSANOGVGMVSAFTVILSKAYPAIGTPIPDITLY 570

QY 40 NIGGDFVATGFCRCVPGAYFFSFTAG-KAPHSLSVWLVNRRDEVQALAFDEORRGA 98  
 DB 571 NKQGHYDPRITGFCRIPLGLYFSYHVHVKGTH--ANVGLYKNGTPVM-XTYDEYIK-GY 626

QY 99 RRAASQASAMLDYDGTWVLRHGAAPHYALGAP--GATFSGYL 140  
 DB 627 LDQASGSAVIDLTENDQVWLQLPNAGSNGLYSEYVHSSFGFLV 671

RESULT 13  
 C1H0QA  
 complement subcomponent C1q chain A precursor [validated] - human  
 N/Alternate names: complement subcomponent C1q alpha chain  
 C/Species: Homo sapiens (man)  
 C/Date: 24-Apr-1984 #sequence\_revision 31-May-1996 #text\_change 08-Dec-2000  
 C/Accession: S14350; A90304; A90315; A903205

R/Sellar, G.C.; Blake, D.J.; Reid, K.B.M.  
 Biochem. J. 274, 481-490, 1991  
 A/Title: Characterization and organization of the genes encoding the A-, B- and C-chain  
 A/Reference number: S14350; MUID:91174759; PMID:1706597  
 C/Accession: S14350  
 A/Status: not compared with conceptual translation  
 A/Molecule type: DNA  
 A/Residues: 1-245 <SEL>  
 R/Reid, K.B.M.  
 Biochem. J. 179, 367-371, 1979  
 A/Title: Complete amino acid sequences of the three collagen-like regions present in su  
 A/Reference number: A90304; MUID:80020137; PMID:486087  
 A/Accession: A90304  
 A/Molecule type: protein  
 A/Residues: 23-96, 'K', 98-102, 'P', 104-130 <REI>  
 R/Reid, K.B.M.; Gagnon, J.; Frampton, J.  
 Biochem. J. 203, 559-569, 1982  
 A/Title: Completion of the amino acid sequences of the A and B chains of subcomponent C  
 A/Reference number: A90315; MUID:82283890; PMID:6981411  
 C/Accession: A90315  
 A/Molecule type: protein  
 A/Residues: 131-171, 'N', 173-177, 'W', 179-239, 'ILRFGSA' <RE2>  
 C/Comment: The first component of complement is a calcium-dependent complex of the thre  
 C/Comment: C1r (enzyme), C1s (proenzyme), and the other eight components of complement.  
 C/Comment: The C1q subcomponent is composed of nine subunits, six of which are disulfid  
 (see PIR:C1HUOC) chain. Equimolar amounts of the A, B, and C chains are found after red  
 C/Genetics:  
 A/Gene: GDB:C1QA  
 A/Cross-references: GDB:119042; OMIM:120550  
 A/Map position: 1p36.3-1p34.1  
 A/Insertions: 55/1  
 C/Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal ho  
 C/Keywords: complement pathway; glycoprotein; heterotrimer; hydroxylysine; hydroxyprolin  
 F:1-22/Domain: signal sequence #status predicted <SIG>  
 F:23-245/Product: complement subcomponent C1q chain A #status experimental <MAT>  
 F:11-109/Domain: collagenous, triple helix #status predicted <COL>  
 F:116-243/Domain: complement C1q carboxyl-terminal homology <C1Q>  
 F:26/Disulfide bonds: Interchain (to chain B-31) #status experimental  
 F:33,48,67,100,103/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F:39,45,54,57,73,85,97/Modified site: 4-hydroxyproline (Pro) #status experimental  
 F:146/Binding site: carboxylate (Asn) (covalent) #status experimental

Query Match 19.7%; Score 147; DB 1; Length 245;  
 Best Local Similarity 30.7%; Pred. No. 1.9e-07;  
 Matches 43; Conservative 21; Mismatches 72; Indels 4; Gaps 3;

QY 4 TTPGSGSELSAASAAATPLTSTSEMAYTFDKYYVIGGDFVATGFCRCVPGAYFFS 63  
 DB 105 SPGNIXQPPRPAFSAIRNPPMGN--VVFDTVTITQEBPYQNHSGRFVCTVPGYYFT 162

QY 64 FTGKAPHSLSVWLVNRRDEVQALAFDEORRGAASQASAMLDYDGTWVLRHGA 123  
 DB 163 FQVLSQWETLSIVSSRGVRRSLGFCDTNKGFLPVVSGVNLQDQSDVWVEKDPK 222

QY 124 P-HYALGAPG-ATFSGYL 141  
 DB 223 KGHLYGSEADSVFSGFLIF 242

RESULT 14  
 CGHU1D  
 collagen alpha 1(X) chain precursor - human  
 N/Alternate names: procollagen alpha 1(X) chain  
 C/Species: Homo sapiens (man)  
 C/Date: 22-Nov-1993 #sequence\_revision 03-Nov-1995 #text\_change 22-Jun-1999  
 C/Accession: S26396; S30086; S18249; A43901; I51870; S21856  
 R/Reichenberger, E.; Beier, F.; LuValle, P.; Olsen, B.R.; von der Mark, K.; Berling, W  
 FEBS Lett. 311, 305-310, 1992  
 A/Title: Genomic organization and full-length cDNA sequence of human collagen X.  
 A/Reference number: S26396; MUID:93012005; PMID:1397333  
 A/Accession: S26396  
 A/Molecule type: DNA

A:Residues: 1-680 <RES1>  
A:Cross-references: EMBL:X66952; EMBL:X72578; EMBL:X72579; EMBL:X72580; GB:S47714; GB:S47715  
A:Appt, S.S.  
submitted to the EMBL Data Library, March 1992  
A:Reference number: S30085  
A:Accession: S30086  
A:Molecule type: DNA  
A:Residues: 'TTPFGWCVWVCLL', 52-680 <APT>  
A:Cross-references: EMBL:X65120; NID:823129  
A:Note: The initial difference is probably due to translation of an intronic sequence  
R:Appt, S.; Mactel, M.G.; Olsen, B.R.  
FEBS Lett. 282, 393-396, 1991  
A:Title: Cloning of human alpha-1(X) collagen DNA and localization of the COL10A1 gene  
A:Reference number: S15826; MUID:91243838; PMID:2037056  
A:Accession: S15826  
A:Molecule type: DNA  
A:Residues: 561-647, 'G', 649-666 <AP2>  
A:Cross-references: EMBL:X58879; NID:930013; PID:CAAA41686.1; PID:930014  
R:Thomas, J.T.; Creswell, C.J.; Raeh, B.; Nicolai, H.; Jones, T.; Solomon, E.; Grant, M.  
Biochem. J. 280, 617-623, 1991  
A:Title: The human collagen X gene. Complete primary translated sequence and chromosomal  
A:Reference number: S18249; MUID:92109659; PMID:1764025  
A:Accession: S18249  
A:Molecule type: DNA  
A:Residues: 1-26, 'T', 28-680 <THO>  
A:Cross-references: EMBL:X60382; NID:930094; PIDN:CAAA2933.1; PID:930095  
A:Note: The sequence from Fig. 3 is inconsistent with that from Fig. 2 in having 532-ATA  
R:Reichenberger, E.; Aigner, T.; von der Mark, K.; Stoss, H.; Betting, W.  
Dev. Biol. 148, 562-572, 1991  
A:Title: In situ hybridization studies on the expression of type X collagen in fetal human  
A:Reference number: A43901; MUID:92077285; PMID:1743401  
A:Accession: A43901  
A:Molecule type: mRNA  
A:Residues: 547-656 <RE2>  
A:Cross-references: GB:M44050; GB:D57494; NID:9339884; PIDN:AA61221.1; PID:9553796  
A:Note: Sequence extracted from NCBI backbone (NCBI:M46012, NCBI:M46014)  
R:Wallis, G.A.; Raeh, B.; Sweetman, W.A.; Thomas, J.T.; Super, M.; Evans, G.; Grant, M.E.  
Am. J. Hum. Genet. 54, 169-178, 1994  
A:Title: Amino acid substitutions of conserved residues in the carboxyl-terminal domain  
pe Schmid.  
A:Reference number: I51870; MUID:94136476; PMID:8304336  
A:Accession: I51870  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 520-597, 'D', 599-680 <WAL>  
A:Cross-references: GB:S66531; NID:9545180; PIDN:AA60615.1; PID:9545181  
A:Note: mutant sequence from patient with metaphyseal chondrodysplasia type Schmid  
A:Note: a second mutant sequence with 614-Pro is also described  
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
ed and subsequently O-glycosylated.  
C:Genetics:  
A:Gene: GDB:COL10A1  
A:Cross-references: GDB:128635; OMIM:120110  
A:Map position: 6q21-6q22  
A:Introns: 52/1  
A:Note: A defect in this gene may cause Schmid metaphyseal chondrodysplasia  
C:Complex: type X collagen may be a homotrimer  
C:Function:  
A:Description: structural component of extracellular fibrous polymer specifically and trimer  
be important for skeletogenesis  
C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
C:Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer; hydroxylysine;  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-680/Product: collagen alpha 1(X) chain #status predicted <MNT>  
F:19-56/Domain: amino-terminal nonhelical #status predicted <MNC2>  
F:57-519/Region: interrupted helical  
F:520-680/Domain: amino-terminal nonhelical #status predicted <NCL>  
F:553-679/Domain: complement C1q carboxyl-terminal homology <CIQ>  
F:617/Binding site: carbohydrate (Aen) (covalent) #status predicted

```

OY      2  GPTFGPSSSELRSAFSA--RTPPL-----EGTSMAT----- 33
Db      511 GPEPPEQAVWPBGFIRAGORPSLSGTRPLSANQGVMPVSAFTVLISKYPAIGTIP 570

OY      34 FDKYVNIIGDFVATGQFRVPGAYFFFTAG-KAPHKSLVWLVNRDEVOALAE 92
Db      571 FDKLVNRQGHYDRTGTIGTQIRGIYFYSVHVVKGN--VWVGLYNGTRVW-YTYDE 627

OY      93 QRPEGARPAASQASAMLDYGDYTWLRHGAH--YALGAPGATFSGYLV 140
Db      628 YTK-GYIDASGSAAIDLTENDQVWLQLPNENGLVSEYVHSFSGFLV 677

RESULT 15
C48150
hibernation-related protein HP-27 precursor - Siberian chipmunk
C/Species: Eutamias sibiricus (Siberian chipmunk)
C/Date: 16-Feb-1994 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C/Accession: C48150; A41752
P/Takamatsu, N.; Ohba, K.; Kondo, J.; Kondo, N.; Shiba, T.
Mol. Cell. Biol. 13, 1516-1521, 1993
A/Title: Hibernation-associated gene regulation of plasma proteins with a collagen-like
A/Reference number: A48150; MUID:93180798; PMID:844393
A/Accession: C48150
A/Molecule type: mRNA
A/Residues: 1-215 <TRK>
A/Cross-references: GB:D12976; NID:G287471; PID:BA00253.1; PID:G287472
A/Note: The source is designated as Tamias amictus in Genbank entry TWSHP27, release
A/Note: sequence extracted from NCBI backbone (NCBIN:125948, NCBIPI:125949)
R/Kondo, N.; Kondo, J.
J. Biol. Chem. 267, 473-478, 1992
A/Title: Identification of novel blood proteins specific for mammalian hibernation.
A/Reference number: A41752; MUID:92112696; PMID:1730610
A/Accession: A41752
A/Status: Preliminary
A/Molecule type: protein
A/Residues: 31-40, 'A', 42-50, 'Q', 52-215 <KON>
C/Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal hc
C/Keywords: glycoprotein; hibernation; plasma
F/1-30/Domain: signal sequence #status predicted <SIG>
F/31-215/Product: hibernation-related protein HP-27 #status experimental <MNT>
F/91-215/Domain: complement C1q carboxyl-terminal homology <CIQ>
F/155/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match      19.4%; Score 144.5; DB 2; Length 215;
Best Local Similarity 32.9%; Pred. No. 3e-07;
Matches 48; Conservative 17; Mismatches 70; Indels 11; Gaps 5;

OY      2  GPTFGP-----SEELSAFSAARTTPLEGSEMAVTPDKYVNIIGDFVATGQFRCV 56
Db      73 GPPEPFGTWNCHSGKTSAP-AVKANELPPAPSGVITKEALHDAQHFDLTGVTTCV 131

OY      57 PGAYFFSTAGKAPHSLSWLVNRNDEVOALAEDEORPGARRAASQAMLDYGDYLV 116
Db      132 PGIVQFGHI-EAVGRAKVSIMRNGTQVMEAEAAQ---DGEYHISGALITQLGMEDRV 187

OY      117 WLRH-GAPHYALGAPGATFSGYLVY 141
Db      188 WLENKLSQTDLERGTQAVFSGFLIH 213

Search completed: January 12, 2004, 08:16:44
Job time : 9.27744 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:09 ; Search time 5.13772 Seconds

(without alignments)  
1308,910 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_17\_159

Sequence: 1 LGPTPGSSSELSAFAAAR.....PHYALGAPATFGSYLVAD 143

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	745	100.0	329	1	COT4_HUMAN
2	196	26.3	285	1	COT2_HUMAN
3	191	25.6	244	1	APM1_HUMAN
4	189	25.4	243	1	COT5_HUMAN
5	188	25.2	278	1	COT6_HUMAN
6	185	24.8	247	1	APM1_MOUSE
7	181.5	24.4	258	1	C1RF_HUMAN
8	181.5	24.4	258	1	C1RF_MOUSE
9	175	23.5	289	1	COT7_HUMAN
10	170.5	22.9	281	1	COT1_HUMAN
11	168.5	22.6	251	1	C1OB_HUMAN
12	167	22.4	419	1	COLE_LERPMA
13	165.5	22.2	255	1	GLIC_MOUSE
14	163.5	21.9	246	1	COT3_HUMAN
15	163.5	21.9	253	1	C1OB_HUMAN
16	160	21.5	245	1	C1OC_HUMAN
17	156.5	21.0	508	1	OTO1_ONCKE
18	156	20.9	245	1	C1OA_MOUSE
19	152.5	20.5	246	1	C1OC_MOUSE
20	151.5	20.3	253	1	C1OB_MOUSE
21	150	20.1	224	1	CERL_MOUSE
22	149	20.0	193	1	CERL_MOUSE
23	149	20.0	193	1	CERB_MOUSE
24	148.5	19.9	674	1	CA1A_CHICK
25	148	19.9	674	1	CA1A_BOVIN
26	147	19.7	245	1	C1OA_HUMAN
27	145	19.5	680	1	C1OA_HUMAN
28	144.5	19.4	215	1	HP27_TAMSI
29	144	19.3	1228	1	ECM_HUMAN
30	141	18.9	743	1	CA18_MOUSE
31	139.5	18.7	635	1	CA28_HUMAN
32	139.5	18.7	744	1	CA18_RABIT
33	138.5	18.6	744	1	CA18_HUMAN

34	136.5	18.3	215	1	HP25_TAMSI	006576	tamias sibi
35	135.5	18.2	170	1	CA28_MOUSE	P25318	mus musculus
36	132	17.7	680	1	CA1A_MOUSE	Q05308	mus musculus
37	126	16.9	201	1	CERL_HUMAN	Q91747	homo sapien
38	113.5	15.2	196	1	HP20_TAMSI	006575	homo sapien
39	74	9.9	181	1	RUNC_TAMSI	Q8Y233	talesonia s
40	74	9.9	272	1	ASRB_SALTY	P26475	salmonella
41	74	9.9	610	1	DNAB_BRECH	Q91CQ5	brevibacilli
42	73	9.8	413	1	IF42_NICPL	P41379	nicotiana p
43	73	9.8	789	1	PRTP_EBV	P25939	epstein-bar
44	69.5	9.3	440	1	PAPA_MOUSE	Q06963	m platelet
45	69	9.3	355	1	IF4X_TOBAC	Q40466	nicotiana t

## ALIGNMENTS

## RESULT 1

COT4\_HUMAN STANDARD; PRT; 329 AA.

AC Q9BXJ3; 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Complement-C1q tumor necrosis factor-related protein 4 precursor.

GN C1QTNF4 OR CTRP4.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Holloway J.L., Lok S.;

RT "Homo sapiens complement-c1q tumor necrosis factor-related protein.";

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

CC - SIMILARITY: Contains 2 C1Q domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: AF329838; AAK17962.1; -

DR Genew: HGNC:14346; C1QTNF4.

DR InterPro: IPR001073; C1q.

DR Pfam: PFO0386; C1q; 2.

DR SMART: SMO0110; C1q; 2.

DR PROSITE: PS01113; C1q; 2.

KW Repeat; signal.

FT SIGNAL 1 16

FT CHAIN 17 329

FT FT 17 329

FT DOMAIN 23 159

FT DOMAIN 170 314

FT SEQUENCE 329 AA; 35265 MW; 331C7DBF26036915 CRC64;

FT POTENTIAL. COMPLEMENT-C1Q TUMOR NECROSIS FACTOR-

FT RELATED PROTEIN 4.

FT C1Q 1.

FT C1Q 2.

FT C1Q 2.

FT C1Q 2.

FT C1Q 2.

FT C1Q 2.

FT C1Q 2.

FT C1Q 2.

FT C1Q 2.

FT C1Q 2.

FT C1Q 2.

FT C1Q 2.

FT C1Q 2.

FT C1Q 2.

FT C1Q 2.

Db 137 HGAPHALGAPGATFSGLYVAD 159

RESULT 2

ID	COT2 HUMAN	STANDARD	PRT	285 AA.
AC	Q9BXJ5			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Complement-c1q tumor necrosis factor-related protein 2 precursor.			
GN	C1QTNF2 OR CTRP2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Piddington C.S., Bishop P.;			
RT	"Homo sapiens complement-c1q tumor necrosis factor-related protein.";			
RL	Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISUS=Muscle;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stephenson M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA	Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,			
RA	Boack S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,			
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Falley J., Helton E., Kerteman M., Madan A., Rodrigues S., Sanchez A.,			
RA	Whiting R.W., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Grimwood J., Schultz J., Myers R.M.,			
RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,			
RA	Scherer A., Schein J.E., Jones S.J.M., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length			
RT	human and mouse cDNA sequences.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
CC	-1- SIMILARITY: Contains 1 C1Q domain.			
CC	-1- SIMILARITY: Contains 1 C1Q domain.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
DR	EMBL, AF329835; AAK17960.1; -			
DR	EMBL, BC016993; AAK1699.1; -			
DR	GeneW, HGNC:14325; C1QTNF2.			
DR	InterPro; IPR001073; C1Q.			
DR	InterPro; IPR000087; Collagen.			
DR	Pfam; PF00386; C1Q; 1.			
DR	Pfam; PF01391; Collagen; 2.			
DR	PRINTS; PRO0007; COMPLEMENTC1Q.			
DR	SMART; SM00110; C1Q; 1.			
DR	PROSITE; PS01113; C1Q; 1.			
KW	Collagen; signal.			
FT	SIGNAL 1 15			
FT	CHAIN 16 285			
FT	POTENTIAL.			
FT	COMPLEMENT-C1Q TUMOR NECROSIS FACTOR-			
FT	RELATED PROTEIN 2.			
FT	COLLAGEN-LIKE.			
FT	DOMAIN 40 141			
FT	DOMAIN 143 285			
FT	C1Q.			
SO	SEQUENCE 285 AA; 29952 MW; 7E31F986BD4EDFA CRC64;			

Query Match 26.3%; Score 196; DB 1; Length 285;

Best Local Similarity 34.2%; Pred. No. 4.5e-13;

Matches 52; Conservative 30; Mismatches 46; Indels 24; Gaps 8;

Db 137 HGAPHALGAPGATFSGLYVAD 159

Query 5 PGP---GSSLRSAFSA--RTTPLEGTSMAVTFKVVYVNIIGDPVATGRCRCRPGA 59

Db 138 PGPSCGSGHTKSAFSAVAVVKSYPRE---RLPTKFDILNKGHYVASSGKVCVPGI 194

Query 60 YPFSFTAGKAPHSLSVLYRNDEVQALAFDEORRGARRAASQSMQLDYGDTYVWL 119

Db 195 YFTYDITLTA-NKHLAIGLVHN-GQYRIRTFD--ANTGNHDVASGRTTLAKGDEVTWLQ 250

Query 120 LHGA-----PHYALGAPGATFSGLYVAD 143

Db 251 IFYSEQNGLFYDPYWT---DSLFTGFLYVAD 278

RESULT 3

ID	APM1 HUMAN	STANDARD	PRT	244 AA.
AC	O15848			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Adiponectin precursor (30 kDa adipocyte complement-related protein)			
DE	(ACRP30) (Adipose most abundant gene transcript 1) (apM-1) (Gelatin-			
DE	binding protein).			
GN	APM1 OR ACRP30 OR GBP28.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISUS=Adipose tissue;			
RX	MEDLINE=96224171; PubMed=8619847;			
RA	Maeda K., Okubo K., Shimomura I., Funahashi T., Matsuzawa Y.,			
RA	Matsubara K.;			
RT	"cDNA cloning and expression of a novel adipose specific collagen-like			
RT	factor. apM1 (Adipose Most abundant Gene transcript 1).";			
RL	Biochem. Biophys. Res. Commun. 221:286-289(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99196984; PubMed=10095105;			
RA	Saito K., Tobe T., Minoshima S., Asakawa S., Sumiya J., Yoda M.,			
RA	Nakano Y., Shimizu N., Tomita M.;			
RT	"Organization of the gene for gelatin-binding protein (GBP28).";			
RL	Gene 229:67-73(1999).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99333693; PubMed=10403784;			
RA	Schaeffler A., Orso E., Palitzsch K.D., Buechler C., Drobnik W.,			
RA	Fuerst A., Schoelmerich J., Schmitz G.;			
RT	"The human apM-1, an adipocyte-specific gene linked to the family of			
RT	TNFs and to genes expressed in activated T cells, is mapped to			
RT	chromosome 1q21.3-q23, a susceptibility locus identified for familial			
RT	combined hyperlipidemia (FCH).";			
RL	Biochem. Biophys. Res. Commun. 260:416-425(1999).			
RN	[4]			
RP	CHARACTERIZATION.			
RX	MEDLINE=20417747; PubMed=10961870;			
RA	Yokota T., Oritani K., Takahashi I., Ishikawa J., Matsuyama A.,			
RA	Ouchi N., Kihara S., Funahashi T., Terner A.J., Tomiyama Y.,			
RA	Matsuzawa Y.;			
RT	"Adiponectin, a new member of the family of soluble defense collagens,			
RT	negatively regulates the growth of myelomonocytic progenitors and the			
RT	functions of macrophages.";			
RL	Blood 96:1723-1732(2000).			
RN	[5]			
RP	CHARACTERIZATION.			
RX	MEDLINE=20440368; PubMed=10982546;			
RA	Ouchi N., Kihara S., Arita Y., Okamoto Y., Maeda K., Kuriyama H.,			

RA Hotta K., Nishida M., Takahashi M., Muraguchi M., Ohmoto Y.,  
 RA Nakamura T., Yamashita S., Funahashi T., Matsuzawa Y.,  
 RT "Adiponectin, an adipocyte-derived plasma protein, inhibits  
 RT endothelial NF-kappaB signaling through a C/EBP-dependent pathway.",  
 RL Circulation 102:1296-1301(2000).  
 RN [6]  
 RP FUNCTION.  
 RX MEDLINE=21372498; PubMed=11479627;  
 RA Yanouchi T., Kamon J., Maki H., Terauchi Y., Kubota N., Hara K.,  
 RA Mori Y., Ide T., Murakami K., Tsuboyama-Kasaoka N., Ezaki O.,  
 RA Akanuma Y., Gavrilova O., Vinson C., Reitman M.L., Kagechika H.,  
 RA Shudo K., Yoda M., Nakano Y., Tobe K., Nagai R., Kimura S., Tomita M.,  
 RA Froguel P., Kadowaki T.;  
 RT "The fat-derived hormone adiponectin reverses insulin resistance  
 RT associated with both lipodystrophy and obesity.",  
 RL Nat. Med. 7:941-946(2001).  
 RN [7]  
 RP VARIANT ADIPONECTIN DEFICIENCY CVS-112.  
 RX MEDLINE=20378830; PubMed=10918532;  
 RA Takahashi M., Arita Y., Yamagata K., Matsukawa Y., Okutomi K.,  
 RA Hori M., Shimomura I., Hotta K., Kuriyama H., Kihara S., Nakamura T.,  
 RA Yamashita S., Funahashi T., Matsuzawa Y.;  
 RT "Genomic structure and mutations in adipose-specific gene,  
 RT Adiponectin",  
 RL Int. J. Obes. Relat. Metab. Disord. 24:861-868(2000).  
 RN [8]  
 RP VARIANTS ARG-84; MET-117; THR-164; SER-221 AND PRO-241.  
 RX MEDLINE=21671103; PubMed=11812766;  
 RA Hara K., Boutin P., Mori Y., Tobe K., Dina C., Yasuda K., Yanouchi T.,  
 RA Orade S., Okada T., Eto K., Kadowaki H., Hagura R., Akanuma Y.,  
 RA Yarakai S., Nagai R., Taniyama M., Matsubara K., Yoda M., Nakano Y.,  
 RA Kimura S., Tomita M., Kimura S., Ito C., Froguel P., Kadowaki T.;  
 RT "Genetic variation in the gene encoding adiponectin is associated with  
 RT an increased risk of type 2 diabetes in the Japanese population.",  
 RL Diabetes 51:536-540(2002).  
 CC -1- FUNCTION: IMPORTANT NEGATIVE REGULATOR IN HEMATOPOIESIS AND IMMUNE  
 CC SYSTEMS; MAY BE INVOLVED IN ENDING INFLAMMATORY RESPONSES THROUGH  
 CC ITS INHIBITORY FUNCTIONS. INHIBITS ENDOTHELIAL NF-KAPPAB SIGNALING  
 CC THROUGH A CAMP-DEPENDENT PATHWAY. INHIBITS TNF-ALPHA-INDUCED  
 CC EXPRESSION OF ENDOTHELIAL ADHESION MOLECULES. INVOLVED IN THE  
 CC CONTROL OF FAT METABOLISM AND INSULIN SENSITIVITY.  
 CC -1- SUBUNIT: HOMODIGOMER (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Synthesized exclusively by adipocytes and  
 CC secreted into plasma.  
 CC -1- DISEASE: Defects in APOL1 are the cause of adiponectin deficiency  
 CC [MIM:605441], resulting in very low concentration of plasma  
 CC adiponectin. Decreased adiponectin plasma levels are associated  
 CC with obesity, insulin resistance, and diabetes type 2.  
 CC -1- PHARMACEUTICAL: Adiponectin might be used in the treatment of  
 CC diabetes type 2 and insulin resistance.  
 CC -1- SIMILARITY: Contains 1 collagenous domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D45371; BAA08227.1; .  
 DR EMBL; AB012165; BAA86716.1; .  
 DR EMBL; AB012164; BAA86716.1; JOINED.  
 DR EMBL; AJ131460; CAB52413.1; .  
 DR EMBL; AJ131461; CAB52413.1; JOINED.  
 DR PIR; JC4708; JC4708.  
 DR MIM; 605441; .  
 DR GO; GO:0006091; P:energy pathways; TAS.  
 DR InterPro; IPR001073; C1q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1q; 1.

DR Pfam; PF01391; Collagen; 1.  
 DR PRINTS; PR00007; COMPLEMENTC1Q.  
 DR ProDom; PD000007; C1q\_helix; 1.  
 DR SMART; SM00110; C1Q; 1.  
 DR PROSITE; PS01113; C1Q; 1.  
 KW Hormone; Collagen; Signal; Repeat; Hydroxylation; Plasma;  
 KW Polymorphism; Disease mutation; Obesity; Diabetes mellitus.  
 FT SIGNAL 1 14  
 FT CHAIN 15 244  
 FT DOMAIN 42 107  
 FT DISUFID 108 244  
 FT DISUFID 36 36  
 FT MOD\_RES 44 44  
 FT MOD\_RES 47 47  
 FT MOD\_RES 53 53  
 FT MOD\_RES 62 62  
 FT MOD\_RES 71 71  
 FT MOD\_RES 76 76  
 FT MOD\_RES 86 86  
 FT MOD\_RES 95 95  
 FT MOD\_RES 104 104  
 FT VARIANT 84 84  
 FT VARIANT 112 112  
 FT VARIANT 117 117  
 FT VARIANT 164 164  
 FT VARIANT 221 221  
 FT VARIANT 241 241  
 SQ SEQUENCE 244 AA; 26414 MW; 64D86C1204B1018 CRC64;  
 Query Match 25.6%; Score 191; DB 1; Length 244;  
 Best Local Similarity 31.8%; Pred. No. 1.2e-12;  
 Matches 47; Conservative 29; Mismatches 54; Indels 18; Gaps 6;  
 QY 5 PGSGSLRGAFAAARTPLEGTSEMAVTFDKYVNIIGSDPVATGQRCRVGAYFFSF 64  
 DB 104 PEGGAYVYRAFSVGLFTYTT-IPNMPIRFTKIFYNQNHYSTGKFRNIGLYFYAY 162  
 QY 65 TAGKAPH-----KSLSMVLVRNDEVOALAFDQRRPFGARRAASQAMQLDGYTVMLR 119  
 DB 163 -----HITVYMDVAVSLFK-KDXKMLFTYDQYQENNDQ-ASGSVTLHLTVGGDQVWLQ 214  
 QY 120 LRGAPH---YALGAFGATFSGLYVAD 143  
 DB 215 VYGEGERNGLYADNDNDSTFTGLVYHD 242  
 RESULT 4  
 ID COTS\_HUMAN STANDARD; PRT; 243 AA.  
 AC Q9BUX0; Q9BUX4;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Complement-c1q tumor necrosis factor-related protein 5 precursor.  
 GN C1QTNF5 OR CTRP5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sheppard P.O., Humes J.M.;  
 RT "Homo sapiens complement-c1q tumor necrosis factor-related protein.",  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RP [2]  
 RP SEQUENCE OF 25-243 FROM N.A.  
 RC TISSUE=Uterus;

```

RA Oltenevaelder B., Obermaier B., Mewes H.-W., Gassenhuber J.,
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 collagenous domain.
CC -1- SIMILARITY: Contains 1 C1Q domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sb.ch/
CC or send an email to license@isb-sb.ch).
CC -----
DR EMBL; AF329841; AAK17965.1; -.
DR EMBL; AL110261; CAB5702.1; -.
DR PIR; T14782; T14782.
DR Genew; HGNC:14344; C1QTNF5.
DR InterPro; IPR001073; C1Q.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF00386; C1q; 1.
DR Pfam; PF01391; Collagen; 1.
DR PRINTS; PR00007; COMPLEMENTC1Q.
DR SMART; SM00110; C1Q; 1.
DR PROSITE; PS0113; C1Q; FALSE_NEG.
KW Collagen; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 243 COMPLEMENT-C1Q TUMOR NECROSIS FACTOR-
FT DOMAIN 30 95 RELATED PROTEIN 5.
FT FT 97 243 COLLAGEN-LIKE.
SQ SEQUENCE 243 AA, 25298 MW, 7CDDA65CDA7EB784 CRC64;

Query Match 25.4%; Score 189; DB 1; Length 243;
Best Local Similarity 35.3%; Pred. No. 26-12;
Matches 53; Conservative 28; Mismatches 57; Indels 12; Gaps 6

QY 2 GPTPGGSEL--RNFSAART-TPLGEGSEMAVTFDKKYNNIGDFVATQGFQCRVPG 58
DB 90 GFTGAGEGCVPRRSFFSAKRESRPVPSDAPLPPDRVLVNEGHYDVAVTGKFTQVNG 149
QY 59 AVEFSTAGAKDHKSISVLTLVNRPDEVOL--AFDEQRPPARRAASGSAMTQLDYGDTV 116
DB 150 VYFPAVHA-TYVRASIQFDLVKNGBSIAFPQFGGMPKPA--SLSGAMVRLPEPDV 205
QY 117 WLRLHGAPE---YALGAPGATFGSYLVYAD 143
DB 206 WVQVGVDYIGIYASIKTSTSGFLVYSD 235

RESULT 5
COT6_HUMAN
ID COT6_HUMAN STANDARD; PRT; 278 AA.
AC 09BX19;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Complement-c1q tumor necrosis factor-related protein 6 precursor.
DE C1QTNF6 OR CTRP6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID:9606;
PI (1)
RP SEQUENCE FROM N.A.
RA Piddington C.S., Sheppard P.O.;
RC "Homo sapiens complement-c1q tumor necrosis factor-related protein.",
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22386257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Dege J.G.,

```

Query Match	Beet Local Similarity	Score 188, DB 1, Length 278, 25.2%, Fred. No. 3e-12, Matches 59, Conservative 18, Mismatches 61, Indels 28, Gaps 7
QY 2 GP--TPGGSS-----ELRS-----AFSAARTPL-EGTSEMAVTFDKYVNVIGG 43		
DB 115 GPGSEPGQSGSKDGMGSPGACQCRFFAFSGRKTAHLSGDFQLLBERVYVNLG 174		
QY 44 DFDVATGQRCRVPCGAYFFSFTAGKAPKHSIVMLVNRNDEVQALAFDEQRPPGARR-AA 102		
DB 175 CFDMATGQFAAPLRIGIYFFSINVSWMYKETYVHIMHQEAVIL---YAQPSERSIMQ 230		
QY 103 SOSAMLTQDYGDTWLR-----HGAPRYALGAPATSSGVIVVND 143		
DB 231 SOSVMLDLAYGDRVVRRLFQKORENAVISNDFITYITSGLIKAE 276		

DE Adiponectin precursor (30 kDa adipocyte complement-related protein)  
 CN APM1 OR ACRP30 OR ADIPOQ.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Adipocyte;  
 RX MEDLINE=96070757; PubMed=7592907;  
 RA Scherer P.E., Williams S., Fogliano M., Baldini G., Lodish H.F.;  
 RT "A novel serum protein similar to C1q, produced exclusively in  
 adipocytes."  
 RL J. Biol. Chem. 270:26746-26749(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Adipocyte;  
 RX MEDLINE=96209999; PubMed=8631877;  
 RA Hu E., Liang P., Spiegelman B.M.;  
 RT "AdipoQ is a novel adipose-specific gene dysregulated in obesity";  
 RL J. Biol. Chem. 271:10697-10703(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=1162543;  
 RA Das K., Lin Y., Widen E., Zhang Y., Scherer P.E.;  
 RT "Chromosomal localization, expression pattern, and promoter analysis  
 of the mouse gene encoding adipocyte-specific secretory protein  
 Acrp30.";  
 RL Biochem. Biophys. Res. Commun. 280:1120-1129(2001).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Heart;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itochi M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Saito T., Okazaki Y., Gojopori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schral L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barin G.,  
 RA Blake J., Botfield D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Guelincich S., Hill D., Hofmann M., Hume D.A., Kanliya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [5]  
 RP FUNCTION.  
 RX MEDLINE=21372498; PubMed=11479627;  
 RA Yamauchi T., Kamon J., Waki H., Teruchi Y., Kubota N., Hara K.,  
 RA Mori Y., Ide T., Murakami K., Tsuboyama-Kasaoka N., Ezaki O.,  
 RA Akanuma Y., Gavrilova O., Vinson C., Reitman M.L., Kagechika H.,  
 RA Shudo K., Yoda M., Nakano Y., Tobe K., Nagai R., Kimura S., Tomita M.,  
 RA Froguel P., Kadowaki T.;  
 RT "The fat-derived hormone adiponectin reverses insulin resistance  
 associated with both lipodystrophy and obesity.";  
 RL Nat. Med. 7:941-946(2001).  
 RN [6]  
 RP FUNCTION.  
 RX MEDLINE=21372499; PubMed=11479628;  
 RA Berg A.H., Combs T.P., Du X., Brownlee M., Scherer P.E.;  
 RT "The adipocyte-secreted protein Acrp30 enhances hepatic insulin  
 action.";  
 RL Nat. Med. 7:947-953(2001).  
 CC -!- FUNCTION: IMPORTANT NEGATIVE REGULATOR IN HEMATOPOIESIS AND IMMUNE

CC SYSTEMS, MAY BE INVOLVED IN ENDING INFLAMMATORY RESPONSES THROUGH  
 CC ITS INHIBITORY FUNCTIONS. INHIBITS ENDOTHELIAL NF-KAPPA-B SIGNALING  
 CC THROUGH A CAMP-DEPENDENT PATHWAY. INHIBITS TNF-ALPHA-INDUCED  
 CC EXPRESSION OF ENDOTHELIAL ADHESION MOLECULES. INVOLVED IN THE  
 CC CONTROL OF FAT METABOLISM AND INSULIN SENSITIVITY.  
 CC -!- SUBUNIT: Homooligomer.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Synthesized exclusively by adipocytes and  
 CC secreted into plasma.  
 CC -!- INDUCTION: DURING HORMONE-INDUCED ADIPOSE DIFFERENTIATION AND  
 CC ACTIVATED BY INSULIN.  
 CC -!- SIMILARITY: Contains 1 collagenous domain.  
 CC -!- SIMILARITY: Contains 1 C1q domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U37222; AAA80543.1; -  
 DR EMBL: U49915; AAB06706.1; -  
 DR EMBL: AF304466; AAK13417.1; -  
 DR EMBL: AK003138; BAB22597.1; -  
 DR PDB: 1C28; 07-SEP-99.  
 DR MGP: MGI:106675; Acrp30.  
 DR GO: GO:0005575; C:extracellular; IDA.  
 DR GO: GO:0005515; F:protein binding activity; IPI.  
 DR GO: GO:0006635; P:faty acid beta-oxidation; IMP.  
 DR InterPro: IPR001073; C1q.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen; 1.  
 DR PRINTS: PR00007; COMPLEMENTC1Q.  
 DR PRODOM: PD000007; C1q\_helix; 1.  
 DR SMART: SM01110; C1Q; 1.  
 DR PROSITE: PS01113; C1Q; 1.  
 KW Hormone; Collagen; Signal; Repeat; Hydroxylation; Plasma;  
 KW Polymorphism; 3D-structure.  
 FT SIGNAL 1 17  
 FT CHAIN 18 247  
 FT DOMAIN 45 110  
 FT DOMAIN 111 247  
 FT DISULFID 39 39  
 FT MOD RES 47 47  
 FT MOD RES 50 50  
 FT MOD RES 56 56  
 FT MOD RES 65 65  
 FT MOD RES 79 79  
 FT MOD RES 98 98  
 FT MOD RES 107 107  
 FT VARIANT 113 113  
 FT CONFLICT 50 50  
 FT CONFLICT 74 74  
 FT CONFLICT 117 117  
 FT CONFLICT 148 148  
 FT CONFLICT 243 243  
 SQ SEQUENCE 247 AA; 26841 MW; 137B687D87398BC4 CRC64;  
 Query Match 24.8%; Score 185; DB 1; Length 247;  
 Best Local Similarity 31.8%; Pred. No. 5; 3e-12;  
 Matches 47; Conservative 29; Mismatches 54; Indels 18; Gaps 7;  
 QY 5 PPGSSSELRSAPSAFTPLEGTSSEMAVTFDKRYVNIIGDFVATGQFCRCRVGAYVFSF 64  
 DB 107 PGRAAAYMRSAPSVGDETRVT-VPNPPIRFTKIFVQNHYGDSGTSKFCNIPGLYFSY 165  
 QY 65 TAGKAPH-----KSLSVMLVRRNDEVQALAPBQKRPGARPARASQAMQLDVGDPVWR 119  
 DB 166 -----HITVYMKDVVSLFK-KDKAVLFTYDYOYKSNVQD-ASGSVLLHLEVGDPVWLD 217

QY 120 LHG-APH---YALGAPGATFSGYLYVAD 143  
 Db 218 VYGDGDHNGLYADNVNDSTFGFLYHYD 245

RESULT 7  
 C1RF\_HUMAN STANDARD; PRT; 258 AA.  
 AC 075973;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE C1q-related factor precursor.  
 GN C1ORF OR CRF.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99097006; PubMed=9878755;  
 RA Berube N.G., Swanson X.H., Berttram M.J., Kittle J.D., Didenko V.,  
 RA Baskin D.S., Smith J.R., Pereira-Smith O.M.;  
 RT "Cloning and characterization of CRF, a novel C1q-related factor,  
 RT expressed in areas of the brain involved in motor function.";  
 RL Brain Res. Mol. Brain Res. 63:233-240(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Rieder M.U., Carrington D.P., Chung M.-W., Lee K.L., Poel C.L., Yi Q.,  
 RA Nickerson D.A.;  
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
 RA Altshul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,  
 RA Hopkin R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stabileton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loguigliano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Blakesley R.W., Touchman A., Young A.C., Shevchenko V., Bouffard G.G.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Buterfield V.S.N., Krzywinski M.J., Skalska U., Smalhus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Maira M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAINSTEM.  
 CC -1- SIMILARITY: Contains 1 collagenous domain.  
 CC -1- SIMILARITY: Contains 1 C1q domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF095154; AAC64186.1; -;  
 CC EMBL: AF410771; AAK95248.1; -;  
 CC EMBL: BC008798; AAH08798.1; -;  
 CC GO: GO:0007626; P: locomotory behavior; NAS.  
 CC InterPro: IPR001073; C1q.  
 CC InterPro: IPR000087; Collagen.

DR Pfam; PF00386; C1q; 1.  
 DR Pfam; PF01391; Collagen; 1.  
 DR PRINTS; PR00007; COMPLEMENTC1Q.  
 DR SMART; SM00110; C1Q; 1.  
 DR PROSITE; PS01113; C1Q; 1.  
 KW Collagen; Signal.  
 FT SIGNAL 1 16  
 FT CHAIN 17 258  
 FT DOMAIN 67 115  
 FT DOMAIN 123 258  
 SO SEQUENCE 258 AA; 26452 MW; 52C51CDF59CAE2BF CRC64;  
 Query Match 24.4%; Score 181.5; DB 1; Length 258;  
 Best local similarity 32.1%; Pred. No. 1,3e-11;  
 Matches 54; Conservative 22; Mismatches 51; Indels 41; Gaps 8;  
 QY 2 GP---TPGPGSSLU-----RSAPSAARTPLEGTSSEMAVTFDKYVNIIGDPFVAT 49  
 Db 106 GPGLPAGAGSGAISTYTTTPRVARYAGLKPHEGYE--VLKFDVVTNIGNNYDAAS 163  
 QY 50 GQFRCRVGAYFFSFTAGKAPHSLSVYLVRNDEVQALAFDEGRPGARRA----- 101  
 Db 164 GKFTCNIPGTYFTY-----H-----VLNRGGDGTSMWA--DLCKNGYRASAIADAD 210  
 QY 102 ----ASQSAMLQDYGDTWLRHGA-PRYALGAPATSGYLYVAD 143  
 Db 211 QNYDASNSVILHLDAGDEVFIKLDGKAGAGNSNKYSTPSGFLIYSD 258

RESULT 8  
 C1RF\_MOUSE STANDARD; PRT; 258 AA.  
 AC 088992;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE C1q-related factor precursor.  
 GN C1ORF OR CRF.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99097006; PubMed=9878755;  
 RA Berube N.G., Swanson X.H., Berttram M.J., Kittle J.D., Didenko V.,  
 RA Baskin D.S., Smith J.R., Pereira-Smith O.M.;  
 RT "Cloning and characterization of CRF, a novel C1q-related factor,  
 RT expressed in areas of the brain involved in motor function.";  
 RL Brain Res. Mol. Brain Res. 63:233-240(1999).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAINSTEM. MORE ABUNDANT IN AREAS  
 CC OF THE NERVOUS SYSTEM INVOLVED IN MOTOR FUNCTION, SUCH AS THE  
 CC PURKINJE CELLS OF THE CEREBELLUM, THE ACCESSORY OLIVARY NUCLEUS,  
 CC THE PONS AND THE RED NUCLEUS.  
 CC -----  
 CC -1- SIMILARITY: Contains 1 collagenous domain.  
 CC -1- SIMILARITY: Contains 1 C1q domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF095155; AAC64187.1; -;  
 CC MGI: 1344400; C1qrf.  
 CC GO: GO:0005737; C: cytoplasm; IDA.  
 CC InterPro: IPR001073; C1q.  
 CC InterPro: IPR000087; Collagen.  
 CC Pfam; PF00386; C1q; 1.  
 CC Pfam; PF01391; Collagen; 1.  
 CC PRINTS; PR00007; COMPLEMENTC1Q.

DR SMART: SM00110; ClQ: 1.  
 DR PROSITE: PS01113; ClQ: 1.  
 KW Collagen; Signal.  
 FT SIGNAL 1 16 POTENTIAL.  
 FT CHAIN 17 258 ClQ-RELATED FACTOR.  
 FT DOMAIN 67 115 COLLAGEN-LIKE.  
 FT DOMAIN 123 258  
 SQ SEQUENCE 258 AA; 26485 MW; F776E2D206B6F763 CRC64;

Query Match 24.4%; Score 181.5; DB 1; Length 258;  
 Best Local Similarity 32.1%; Pred. No. 1,3e-11;  
 Matches 54; Conservative 22; Mismatches 51; Indels 41; Gaps 8;

QY 2 GP--TTPGSSSEL-----RSAPSAARTTPLEGTSMAVTFDKRYVNIQGDPEVAT 49  
 DB 106 GPGPLPGSGSGGSIATYTTTTPRVAVAGLKNPHEGYE--VLKFDVVTNIGNNYDAAS 163  
 QY 50 GQRCRVPAGVFPFSTAGKAPHSLSVMLVRNDEVQALAFDQRRGARRA----- 101  
 DB 164 GKFTCNIPGTYFTY-----H-----VLNRGGDGTSMWA--DLCKNGQVRASAIADAD 210  
 QY 102 ----ASQAMQLQDYGDTVWLRHGA-PHYALGAPGATPGSYLVYAD 143  
 DB 211 QNVYASNSVYLHLHDAGDEVFIKLDGSKAGGNSKSTSTSGFTIYSD 258

RESULT 9  
 COT7 HUMAN STANDARD; PRT; 289 AA.  
 ID COT7 HUMAN  
 AC 09BKJ2;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Complement-c1q tumor necrosis factor-related protein 7 precursor.  
 GN C1QTNF7 OR CTRP7.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_Taxid=9606;  
 RP [1]  
 RA SEQUENCE FROM N.A.  
 RT Piddington C.S., Sheppard P.O., Bishop P., Lasser G.W.;  
 RL "Homo sapiens complement-c1q tumor necrosis factor-related protein.",  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.F., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.U., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Miliang S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Murry D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butcherfield Y.S.N., Krzywiński M.I., Skalska U., Smalls D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Merra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- SIMILARITY: Contains 1 collagenous domain.  
 CC -1- SIMILARITY: Contains 1 ClQ domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF329637; AK17963.1; -  
 DR EMBL: BC022187; AAH22187.1; -  
 DR Genew; HGNC:14342; C1QTNF7.  
 DR InterPro: IPR001073; C1QTNF7.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF00386; Clq; 1.  
 DR Pfam: PF01391; Collagen; 2.  
 DR PRINTS: PR00007; COMPLEMENTC1Q.  
 DR SMART: SM00110; ClQ; 1.  
 DR PROSITE: PS01113; ClQ; 1.  
 KW Collagen; Signal.  
 FT SIGNAL 1 16 POTENTIAL.  
 FT CHAIN 17 289 COMPLEMENT-C1Q TUMOR NECROSIS FACTOR-  
 FT DOMAIN 38 139 RELATED PROTEIN 7.  
 FT DOMAIN 141 276 COLLAGEN-LIKE.  
 SQ SEQUENCE 289 AA; 30683 MW; A61609FF86D26946 CRC64;

Query Match 23.5%; Score 175; DB 1; Length 289;  
 Best Local Similarity 33.5%; Pred. No. 7e-11;  
 Matches 56; Conservative 21; Mismatches 58; Indels 32; Gaps 8;

QY 1 LGPTGEP-----GSELRSAPSAARTT--PLEGTSMAVTFDKRYVN 40  
 DB 118 IGP-PGPKGDRGEQDPPGLPGVCRCGSYLVKSAFSVGTITSYEE--RLPIIFNVLVN 173  
 QY 41 ICGDPVATGQRCRVPAGVFPFSTAGKAPHSLSVMLVRNDEVQALAFDQRRGARR 100  
 DB 174 ESEHNPAITGKFTCAFPGLTYFTSYDITLA-NHGLAIGLVAN-GQRIKTFD--ANTGNHD 229

QY 101 AASQAMQLQDYGDTVWLRHGA-PHYAL---GAPGATPGSYLVYAD 143  
 DB 230 VASGSTVYILQPEDEVLEIFFTDQNGLPSPDGMDSLPSGFLVYD 276

RESULT 10  
 COT1 HUMAN STANDARD; PRT; 281 AA.  
 ID COT1 HUMAN  
 AC 09BKJ1; O96NF2; O9GZ84;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Complement-c1q tumor necrosis factor-related protein 1 precursor  
 DE (G protein coupled receptor interacting protein) (GIP).  
 GN C1QTNF1 OR CTRP1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_Taxid=9606;  
 RP [1]  
 RA SEQUENCE FROM N.A.  
 RT Innamorti G., Le Gouill C., Whang I., Birnbaumer M.;  
 RL "GIP, a putative GPCR interacting protein.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

122 GAPHALGAPG-----ATPSGYLV 140  
| : :  
| : :  
| : :  
| : :  
253 GERENNAIFSEELDYITTPSGYLV 275

RESULT 11

ID	C1QB_HUMAN	STANDARD;	PRT;	251 AA.
AC	P02746; Q96H17;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Complement C1q subcomponent, B chain precursor.			
GN	C1QB.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE=86076906; PubMed=3000358;			
RX	Reid K.B.M.;			
RT	"Molecular cloning and characterization of the complementary DNA and gene coding for the B-chain of subcomponent C1q of the human complement system.";			
RL	Biochem. J. 231:729-735(1985).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=22388257; PubMed=12479312;			
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hatsh F., Diatchenko L., Marziska K., Farmer A.A., Rubin G.M., Hong L., Stjepanec M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Udell T.B., Toshiyuki S., Abramson R.D., Mullaly S.J., Rana S.S., Loguallano N.A., Peters G.U., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Wozley K.C., Hale S., Garcia A.M., Gay L.J., Huliyk S.W., Villalon D.K., Muzy D.M., Sodergren E.J., Lu X., Gibbs R.A., Paley J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butleriez J.S.N., Krzywinski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences".			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16699-16903(2002).			
RN	[3]			
RP	SEQUENCE OF 26-133.			
RX	MEDLINE=80020137; PubMed=486087;			
RA	Reid K.B.M.;			
RT	"Complete amino acid sequences of the three collagen-like regions present in subcomponent C1q of the first component of human complement.";			
RL	Biochem. J. 179:367-371(1979).			
RN	[4]			
RP	SEQUENCE OF 26-193.			
RX	MEDLINE=79041552; PubMed=708376;			
RA	Reid K.B.M., Thompson E.O.P.;			
RT	"Amino acid sequence of the N-terminal 108 amino acid residues of the B chain of subcomponent C1q of the first component of human complement.";			
RL	Biochem. J. 173:863-866(1978).			
RN	[5]			
RP	SEQUENCE OF 134-251.			
RX	MEDLINE=82283890; PubMed=6981411;			
RA	Reid K.B.M., Gagnon J., Frampton J.;			
RT	"Completion of the amino acid sequences of the A and B chains of subcomponent C1q of the first component of human complement.";			
RL	Biochem. J. 203:559-569(1982).			



RN [6]  
 RP SEQUENCE OF 224-251 FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=85038855; PubMed=6208566;  
 RA Reid K.B.M., Bentley D.R., Wood K.J.;  
 RT "Cloning and characterization of the complementary DNA for the B  
 chain of normal human serum C1q";  
 RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 306:345-354 (1984).  
 RP REVIEW OF C1Q DEFICIENCY.  
 RX MEDLINE=98450587; PubMed=9777412;  
 RA Petry F.;  
 RT "Molecular basis of hereditary C1q deficiency";  
 RL Immunobiology 199:286-294 (1998).  
 CC -1- FUNCTION: C1Q ASSOCIATES WITH THE PROENZYMES C1R AND C1S TO YIELD  
 C1, THE FIRST COMPONENT OF THE SERUM COMPLEMENT SYSTEM. THE  
 COLLAEN-LIKE REGIONS OF C1Q INTERACT WITH THE CA(2+)-DEPENDENT  
 C1R(2)C1S(2) PROENZYME COMPLEX, AND EFFICIENT ACTIVATION OF C1  
 TAKES PLACE ON INTERACTION OF THE GLOBULAR HEADS OF C1Q WITH THE  
 FC REGIONS OF IGG OR IGM ANTIBODY PRESENT IN IMMUNE COMPLEXES.  
 CC -1- SUBUNIT: C1 IS A CALCIUM-DEPENDENT TRIMOLECULAR COMPLEX OF C1Q, R  
 AND S IN THE MOLAR RATION OF 1:2:2. C1Q SUBCOMPONENT IS COMPOSED  
 OF NINE SUBUNITS, SIX OF WHICH ARE DISULFIDE-LINKED DIMERS OF THE  
 A AND B CHAINS, AND THREE OF WHICH ARE DISULFIDE-LINKED DIMERS OF  
 THE C CHAIN.  
 CC -1- PTM: O-linked glycans consist of Glc-Gal disaccharides bound to  
 the oxygen atom of post-translationally added hydroxyl groups.  
 CC -1- DISEASE: Defects in C1QB are a cause of C1Q deficiency  
 [MM:120570]. It is a rare genetic disorder which is associated  
 with recurrent infections and a high prevalence of lupus  
 erythematosus-like symptoms. It is characterized by a loss of  
 activation of the complement classical pathway.  
 CC -1- SIMILARITY: Contains 1 C1Q domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X03084; CAA26880.1; -  
 DR EMBL; BC008983; AA008983.1; AUT\_INIT.  
 DR EMBL; M36278; AAC1692.1; -  
 DR Genew; HGNC:1242; C1QB.  
 DR MIM: 120570; -  
 DR GO; GO:0005602; C:complement component C1q complex; TAS.  
 DR GO; GO:0003811; F:complement activity activity; TAS.  
 DR InterPro; IPR001073; C1q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1q; 1.  
 DR Pfam; PF01391; Collagen; 1.  
 DR PRINTS; PRO0007; COMPLEMENTC1Q.  
 DR SMART; SM0110; C1Q; 1.  
 DR PROSITE; PS01113; C1Q; 1.  
 KW Complement pathway; Plasma; Hydroxylation; Glycoprotein; Collagen;  
 Repeat; Signal; Disease mutation; Pyrrolidone carboxylic acid.  
 FT SIGNAL 1 25  
 FT CHAIN 26 251  
 FT DOMAIN 29 112  
 FT MOD\_RES 113 251  
 FT MOD\_RES 26 26  
 FT DISULFID 29 29  
 FT MOD\_RES 33 33  
 FT MOD\_RES 33 33  
 FT MOD\_RES 36 36  
 FT MOD\_RES 39 39  
 FT MOD\_RES 42 42  
 FT MOD\_RES 51 51  
 FT MOD\_RES 54 54  
 FT MOD\_RES 57 57  
 FT CARBOHYD 57 57  
 O-LINKED (GAL. . .)

FT MOD\_RES 60 60  
 FT CARBOHYD 60 60  
 FT MOD\_RES 63 63  
 FT MOD\_RES 63 63  
 FT MOD\_RES 75 75  
 FT MOD\_RES 81 81  
 FT MOD\_RES 84 84  
 FT MOD\_RES 90 90  
 FT MOD\_RES 96 96  
 FT MOD\_RES 96 96  
 FT CARBOHYD 96 96  
 FT MOD\_RES 99 99  
 FT MOD\_RES 102 102  
 FT MOD\_RES 105 105  
 FT MOD\_RES 108 108  
 FT MOD\_RES 108 108  
 FT CARBOHYD 108 108  
 FT VARIANT 40 40  
 FT CONFLICT 26 26  
 FT CONFLICT 83 83  
 FT CONFLICT 98 98  
 FT SEQUENCE 251 AA; 26459 MW; 78C5752E267A0EF7 CRC64;  
 Query Match 22.6%; Score 168.5; DB 1; Length 251;  
 Best Local Similarity 27.7%; Pred. No. 2.8e-10;  
 Matches 44; Conservative 34; Mismatches 62; Indels 19; Gaps 4;  
 QY 1 LGPTPGFS-----SELSAFSAARTTPELSTSEMAVTFDKVYVNIIGDFD 46  
 DB 93 MGPGGAGGADGAPGKSGSDYKATQIARSARTIIVPRLRDOTIFDHVITMNNNVE 152  
 QY 47 VAIQGFRCRPGAFYFSTGKAPHSLSVLYRNDDEVQALAFDEQRREGARRAASQSA 106  
 DB 153 PRSGKFKCKVPGYGFYFTH--SSRGMLCVNMRGRRAQKVYTFCDYAVNTQVTTGGM 210  
 QY 107 MLQIDYGDVTWLRHGAHYALGAPAT--FSGLYVAD 143  
 DB 211 VLKLEQENVFLQATD-KNSLGMEGANSIFSGFLFPD 248  
 RESULT 12  
 COLE LEPMA STANDARD; PRT; 419 AA.  
 AC P98085; O91080;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Inner ear-specific collagen precursor (Saccular collagen).  
 OS Lepomis macrochirus (Bluegill)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorphia; Acanthopterygii; Perciformes; Percoidae;  
 OC Centrarchidae; Lepomis.  
 ON NCB1\_TaxID=13106;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95167486; PubMed=7863311;  
 RA Davis J.G., Oberholzer J.C., Burns F.R., Greene M.I.;  
 RT "Molecular cloning and characterization of an inner ear-specific  
 structural protein";  
 RL Science 267:1031-1034 (1995).  
 RN [2]  
 RP CONCEPTUAL TRANSLATION.  
 RA Gibson T.;  
 RL Submitted (MAR-1995) to the SWISS-PROT data bank.  
 CC -1- FUNCTION: FORMS A MICROSTRUCTURAL MATRIX WITHIN THE OTOLITHIC  
 MEMBRANE (PROBABLY).  
 CC -1- TISSUE SPECIFICITY: SPECIALIZED SECRETORY SUPPORTING CELLS AT THE  
 OUTER PERIMETER OF THE SACCULAR EPITHELIUM.  
 CC -1- SIMILARITY: CONTAINS THE SACCULAR C1Q DOMAIN.  
 CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT HAD TO BE  
 INTRODUCED FOR POSITIONS 391-419 SO AS TO MAXIMIZE THE SIMILARITY  
 WITH OTHER SHORT-CHAIN COLLAGENS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL: U17431; AAA69978.1; ALT\_FRAME.  
 DR InterPro: IPR001073; Clq.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF00386; Clq; 1.  
 DR Pfam: PF01391; Collagen; 3.  
 DR PRINTS: PR00007; COMPLEMENTC1Q.  
 DR PRODOM: PD000007; Clq\_helix; 2.  
 DR SMART: SMO0110; Clq; 1.  
 DR PROSITE: PS01113; Clq; 1.  
 DR Extracellular matrix; Repeat; Collagen; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 419  
 FT DOMAIN 20 57  
 FT DOMAIN 58 274  
 FT DOMAIN 275 419  
 FT DOMAIN 272 419  
 FT CARBOHYD 37 37  
 FT CARBOHYD 320 320  
 SQ SEQUENCE 419 AA; 43634 MW; 570CDB9675FC0F39 CRC64;  
 Query Match 22.4%; Score 167; DB 1; Length 419;  
 Best Local Similarity 32.9%; Pred. No. 7.3e-10;  
 Matches 52; Conservative 20; Mismatches 56; Indels 30; Gaps 8;

QY 2 GPTPGGSS--ELRSAPSA---ARTTPLEGTESEMAVTFDKYVYVIGGDFVATGQFRCR 55  
 DB 266 GKPGKPGSVGEIRSAFSGVGLFPRSFP--PPSLPVKEDKVFYNGEGHMDPTLKNFNT 322  
 QY 56 VEGAVFESF--TAGCAPKSL---SWLVNRDEVQALAPDEORRPPGARRAASQAMIQ 109  
 DB 323 YPGVYLFYSYHITVRNRPVPAALVAVGVKRLTRDSLYGGDIDIQ-----ASNLALH 373  
 QY 110 LDYGDVTWMLR---HGAPVHALGAPGATFSGYLVYAD 143  
 DB 374 LTDGQVWLETLRDMNGX--YSSSEDDTFSGFLLPD 409

RESULT 13  
 GLIC MOUSE STANDARD; PRT; 255 AA.

AC Q9ESN4;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Gliacolin precursor (Clq-like protein).  
 GN Clq.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20428709; PubMed=10862616;  
 RA Koide T., Aso A., Yoshizumi T., Nagata K.;  
 RT "Conformational requirements of collagenous peptides for recognition  
 RT by the chaperone protein HSP47.";  
 RL J. Biol. Chem. 275:27957-27963(2000).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN GLIAL CELLS.  
 CC -1- SIMILARITY: Contains 1 collagenous domain.  
 CC -1- SIMILARITY: Contains 1 Clq domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL: AB04560; BAB15806.1; -  
 DR MED: MG1:2387350; Clq1.  
 DR GO: GO:0005515; F-protein binding activity; IPI.  
 DR InterPro: IPR001073; Clq.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF00386; Clq; 1.  
 DR Pfam: PF01391; Collagen; 1.  
 DR PRINTS: PR00007; COMPLEMENTC1Q.  
 DR SMART: SMO0110; Clq; 1.  
 DR PROSITE: PS01113; Clq; 1.  
 KW Collagen; Signal.  
 FT SIGNAL 1 20  
 FT CHAIN 21 255  
 FT DOMAIN 61 111  
 FT DOMAIN 120 255  
 SQ SEQUENCE 255 AA; 26687 MW; 529FBAF4B219BCL CRC64;

Query Match 22.4%; Score 165.5; DB 1; Length 255;  
 Best Local Similarity 30.4%; Pred. No. 5.9e-10;  
 Matches 48; Conservative 23; Mismatches 66; Indels 21; Gaps 5;

QY 2 GPTPGG-----SSELRSAFSAARTTPLEGTESEMAVTFDKYVYVIGGDFVAT 49  
 DB 103 GPPGAPGIMAGAIASATYSTVPKIAFYAGLKRGEGYE--VLKFDVVTNLTGNHYDPTT 160  
 QY 50 GQFRCVPGAVFESF--TIGKAPKSLSWLVNRDEVQALAPDEORRPPGARRAASQSA 106  
 DB 161 GKFTCSIPGIFYFYHYVLMRGDGTSMWADICKNNVPAASAIADADON--YYASNSV 217  
 QY 107 MLQLDYGDVTWMLRHLGA--PHYALGAPGATFSGYLVYAD 143  
 DB 218 VLHPEPDGVYIKLDGKGAGGNNKXSTFSGFLIYAD 255

RESULT 14  
 COT3 HUMAN STANDARD; PRT; 246 AA.

AC Q9BX74; Q96KY1;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Complement-c1q tumor necrosis factor-related protein 3 precursor  
 DE (Secretory protein COR326).  
 GN C1QTNF3 OR CTRP3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Piddington C.S., Bishop P.;  
 RT "Homo sapiens complement-c1q tumor necrosis factor-related protein";  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Maeda T., Hayashi A., Saito T.;  
 RT "Molecular cloning, chromosomal localization, and genomic structure of  
 RT the human COR326 gene";  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 121-246 FROM N.A.  
 RC TISSUE-Placenta; PubMed=12477932;  
 RX MEDLINE=22388257;  
 RA Straussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shermen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo W.F., Caavaant T.L., Scheetz T.F.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loguella N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gumatratte P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmitz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.L., Skalska U., Smallus D.E.,  
 RA Schercher A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- SIMILARITY: Contains 1 C1Q domain.  
 CC -1- SIMILARITY: Contains 1 C1Q domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF329837; AAK17961.1; -  
 DR EMBL: AF326976; AAK70344.1; -  
 DR EMBL: BC016021; AAH16021.1; -  
 DR Genew: HGNC:14326; C1QTF3.  
 DR InterPro: IPR001073; C1Q.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF00386; C1Q; 1.  
 DR Pfam: PF01391; Collagen; 1.  
 DR SMART: SM00110; C1Q; 1.  
 DR PROSITE: PS01113; C1Q; 1.  
 DR KW Collagen; Signal.  
 FT CHAIN 1 22 POTENTIAL.  
 FT SIGNAL 23 246 COMPLEMENT-C1Q TUMOR NECROSIS FACTOR-  
 FT DOMAIN 51 113 RELATED PROTEIN 3.  
 FT DOMAIN 114 246 COLLAGEN-LIKE.  
 FT CONFLICT 214 214 C1Q.  
 SQ SEQUENCE 246 AA; 26994 MW; C589B6C3A735ED29 CRC64;  
 K -> E (IN REF. 3).  
 Query Match 21.9%; Score 163.5; DB 1; Length 246;  
 Best Local Similarity 35.2%; Pred. No. 9.1e-10;  
 Matches 51; Conservative 24; Mismatches 49; Indels 21; Gaps 8;  
 QY 7 PG-SSELSARFARTTPLEGTSMAVYFDKYVYVNIIGDPPVATGQFCRCRPGAYFBSFT 65  
 DB 110 PGIPPELQIAFMASLATHFSNONS-GILFSSVEVETNIGFVMTGRCFAPVSGVYFFTS 168  
 QY 66 AGKAPHSKLS---VNLVNRDEVOALAFDEQRPGARRASQASAMLOLDYGDYWLRL-H 121  
 DB 169 MMK--HEVEVEVYVYLVNNGTVFSMYSYEMK--GKSDTSSNNAVLYKLAKDEWLVKGN 224  
 QY 122 GAPHYALGAPG---ATFSGYLVY 141  
 DB 225 GALT-----GDHORFTFAGFLF 243  
 RESULT 15  
 C1Q-RAT STANDARD; PRT; 253 AA.  
 AC P31721;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Complement C1q subcomponent, B chain precursor.  
 GN C1Q.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Spleen;  
 RX MEDLINE=96062319; PubMed=7594503;  
 RA Schwable W., Schaefer M.K.-H., Petty F., Fink T., Knebel D.,  
 RA Welhe E., Loos M.,  
 RT "Pollicular dendritic cells, interdigitating cells, and cells of the  
 RT monocyte-macrophage lineage are the C1q-producing sources in the  
 RT spleen. Identification of specific cell types by in situ  
 RT hybridization and immunohistochemical analysis.";  
 RL J. Immunol. 155:4971-4978(1995).  
 RP SEQUENCE OF 71-79 AND 141-146.  
 RX MEDLINE=93218657; PubMed=8464426;  
 RA Wing M.G., Seilly D.J., Bridgman D.J., Harrison R.A.,  
 RT "Rapid isolation and biochemical characterization of rat C1 and C1q,"  
 RL Mol. Immunol. 30:433-440(1993).  
 CC -1- FUNCTION: C1Q ASSOCIATES WITH THE PROENZYMES C1R AND C1S TO YIELD  
 CC C1, THE FIRST COMPONENT OF THE SERUM COMPLEMENT SYSTEM. THE  
 CC COLLAGEN-LIKE REGIONS OF C1Q INTERACT WITH THE CA(2+)-DEPENDENT  
 CC C1R(2)C1S(2) PROENZYME COMPLEX, AND EFFICIENT ACTIVATION OF C1  
 CC TAKES PLACE ON INTERACTION OF THE GLOBULAR HEADS OF C1Q WITH THE  
 CC FC REGIONS OF IGG OR IGM ANTIBODY PRESENT IN IMMUNE COMPLEXES.  
 CC -1- SUBUNIT: C1 IS A CALCIUM-DEPENDENT TRIMOLECULAR COMPLEX OF C1Q, R  
 CC AND S IN THE MOLAR RATION OF 1:2:2. C1Q SUBCOMPONENT IS COMPOSED  
 CC OF NINE SUBUNITS, SIX OF WHICH ARE DISULFIDE-LINKED DIMERS OF THE  
 CC A AND B CHAINS, AND THREE OF WHICH ARE DISULFIDE-LINKED DIMERS OF  
 CC THE C CHAIN. IN ADDITION TO THE MAJOR A:B AND C:C DIMER BANDS,  
 CC RAT, UNLIKE HUMAN C1Q, CONTAINED MINOR DIMER SPECIES.  
 CC -1- SIMILARITY: Contains 1 collagenous domain.  
 CC -1- SIMILARITY: Contains 1 C1Q domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X71127; CAA50440.1; -  
 DR PIR: S49158; S49158.  
 DR InterPro: IPR001073; C1Q.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF00386; C1q; 1.  
 DR Pfam: PF01391; Collagen; 1.  
 DR PRINTS: PR00007; COMPLEMENTC1Q.  
 DR SMART: SM00110; C1Q; 1.  
 DR PROSITE: PS01113; C1Q; 1.  
 DR KW Complement pathway; Plasma; Hydroxylation; Glycoprotein; Collagen;  
 KM Repeat; Signal.  
 FT CHAIN 1 25  
 FT SIGNAL 26 253  
 FT DOMAIN 29 112  
 FT DOMAIN 113 253  
 FT DISULFID 29 29  
 FT MOD\_RES 33 33  
 FT MOD\_RES 33 33  
 FT MOD\_RES 36 36  
 FT MOD\_RES 39 39  
 FT MOD\_RES 39 39  
 FT MOD\_RES 51 51  
 FT MOD\_RES 51 51  
 FT MOD\_RES 54 54  
 FT MOD\_RES 57 57  
 FT MOD\_RES 60 60  
 FT MOD\_RES 63 63  
 FT MOD\_RES 63 63  
 FT MOD\_RES 75 75  
 FT MOD\_RES 81 81  
 FT MOD\_RES 81 81  
 FT MOD\_RES 84 84  
 FT MOD\_RES 90 90  
 FT MOD\_RES 96 96  
 FT MOD\_RES 96 96  
 FT MOD\_RES 99 99  
 SQ SEQUENCE 253 AA; 26589 MW; 1CB40622571BFC9B CRC64;  
 Query Match 21.9%; Score 163.5; DB 1; Length 253;  
 QY 1 INTERCHAIN (WITH C-26 IN CHAIN A).  
 DB 110 INTERCHAIN (WITH C-26 IN CHAIN A).  
 QY 66 AGKAPHSKLS---VNLVNRDEVOALAFDEQRPGARRASQASAMLOLDYGDYWLRL-H 121  
 DB 169 MMK--HEVEVEVYVYLVNNGTVFSMYSYEMK--GKSDTSSNNAVLYKLAKDEWLVKGN 224  
 QY 122 GAPHYALGAPG---ATFSGYLVY 141  
 DB 225 GALT-----GDHORFTFAGFLF 243  
 RESULT 15  
 C1Q-RAT STANDARD; PRT; 253 AA.  
 AC P31721;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Complement C1q subcomponent, B chain precursor.  
 GN C1Q.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;

Best Local Similarity 29.2%; Pred. No. 9.4e-10;  
Matches 45; Conservative 36; Mismatches 58; Indels 15; Gaps 7;

```

QY 2 GPTGCP-----GSELRS----AFSAARTTPLEGTSEMAVTFDKYVYVNIIGDPDFVATGQF 52
Db 100 GP-PGPRGPKGSGDYKATQKVAFSALRTVNSALRPNOAIRPEKVTITVNDNYEPRSGKF 158
QY 53 RCRVPGAYFFSFTAGKAPHSLSVMLVNNRDEVQ-ALAFDEQRPRPGARRAASOSAMLOLD 111
Db 159 TCKVPGLYFTFYHASRGNLCVNIVRGRDRDRMOKVLTFCDYAQ-NTFQVTTGGVVLKLE 217
QY 112 YGDTYMLRLHGAAPHYALGAPGAT--FSGYLYAD 143
Db 218 QBEVVHLQATD-KNSLLGVEGANSIFTGFLFPD 250

```

Search completed: January 12, 2004, 08:12:58  
Job time : 6.13772 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:10 ; Search time 19.6946 Seconds

(Without alignments)  
1873.686 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_17\_159

Sequence: 1 LGPTPGSGSELRSASFSAAR.....PHYALGAPGATFSGLYLVAD 143

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP\_TREMBL\_23:\*  
2: SP\_Bacteria:\*  
3: SP\_Fungi:\*  
4: SP\_Human:\*  
5: SP\_Invertebrate:\*  
6: SP\_Mammal:\*  
7: SP\_Mhc:\*  
8: SP\_Organelle:\*  
9: SP\_Phage:\*  
10: SP\_Plant:\*  
11: SP\_Rodent:\*  
12: SP\_Virus:\*  
13: SP\_Vertebrate:\*  
14: SP\_Unclassified:\*  
15: SP\_Virus:\*  
16: SP\_Bacteriophage:\*  
17: SP\_Archaeop:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	737	98.9	329	4 Q81V25	Q81V25 homo sapien
2	706.5	94.8	326	11 Q8R066	Q8R066 mus musculu
3	355	47.7	205	11 Q9D0W2	Q9D0W2 mus musculu
4	349	46.8	205	11 Q9DCB6	Q9DCB6 mus musculu
5	222	29.8	158	4 Q9H667	Q9H667 homo sapien
6	222	29.8	1077	4 Q8TE71	Q8TE71 homo sapien
7	216	29.0	158	11 Q8K110	Q8K110 mus musculu
8	200	26.8	294	11 Q9D8U4	Q9D8U4 mus musculu
9	192	25.8	243	6 Q95JD7	Q95JD7 macaca mula
10	189	25.4	243	4 Q8NEP2	Q8NEP2 homo sapien
11	185	24.8	247	11 Q8BRW2	Q8BRW2 mus musculu
12	181.5	24.4	243	11 Q8R002	Q8R002 mus musculu
13	180.5	24.2	264	11 Q8BKRO	Q8BKRO mus musculu
14	180	24.2	244	11 Q8K3R4	Q8K3R4 rattus norv
15	179	24.0	240	6 Q95MQ4	Q95MQ4 bos taurus
16	178.5	24.0	243	11 Q8K479	Q8K479 mus musculu

17	171.5	23.0	312	11 Q8CHX9	Q8CHX9 mus musculu
18	171	23.0	287	11 Q8CFR0	Q8CFR0 mus musculu
19	165.5	22.2	194	6 Q95J95	Q95J95 canis fam1
20	164.5	22.1	120	11 Q8R1Z2	Q8R1Z2 mus musculu
21	162.5	21.8	246	11 Q9E330	Q9E330 mus musculu
22	162.5	21.8	289	11 Q8BYD7	Q8BYD7 mus musculu
23	160.5	21.5	182	11 Q8R1P2	Q8R1P2 mus musculu
24	160.5	21.5	281	11 Q9QXP7	Q9QXP7 mus musculu
25	156	20.9	245	11 Q9DCM6	Q9DCM6 mus musculu
26	150	20.1	224	4 Q81UX8	Q81UX8 homo sapien
27	150	20.1	224	11 Q8BGU2	Q8BGU2 mus musculu
28	146.5	19.7	197	11 Q9JHC0	Q9JHC0 mus musculu
29	144.5	19.4	333	4 Q81U04	Q81U04 homo sapien
30	142.5	18.9	173	6 Q62789	Q62789 sus scrofa
31	141	18.1	744	11 Q921S8	Q921S8 mus musculu
32	141	18.9	744	11 Q8BGL6	Q8BGL6 mus musculu
33	139	18.7	675	6 Q9N178	Q9N178 mus musculu
34	135.5	18.2	705	4 Q8T5U5	Q8T5U5 sus scrofa
35	134.5	18.1	213	5 P83425	P83425 mytilus edu
36	132.5	17.8	195	11 Q8BZS3	Q8BZS3 mus musculu
37	131	17.6	185	5 Q9GQV4	Q9GQV4 strongyloce
38	130	17.4	222	13 Q8J126	Q8J126 salvelinus
39	127.5	17.1	102	6 Q95MB2	Q95MB2 equus cabal
40	127.5	17.1	225	11 Q9Z1K4	Q9Z1K4 rattus norv
41	127.5	17.1	1017	11 Q99K41	Q99K41 mus musculu
42	127	17.0	198	11 Q8BMFO	Q8BMFO mus musculu
43	126	16.9	198	11 Q8BME9	Q8BME9 mus musculu
44	121.5	16.3	347	4 Q96IH6	Q96IH6 homo sapien
45	121.5	16.3	583	4 Q96G58	Q96G58 homo sapien

## ALIGNMENTS

RESULT 1  
Q81V25 PRELIMINARY; PRT; 329 AA.  
AC Q81V25;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Similar to C1g and tumor necrosis factor related protein 4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC035628; AAH35628.1; -; 16064DA8182A6732 CRC64;  
SQ SEQUENCE 329 AA; 35256 MW; 16064DA8182A6732 CRC64;

Query Match 98.9%; Score 737; DB 4; Length 329;  
Best Local Similarity 99.3%; Pred. No. 1.3e-68;  
Matches 142; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	LGPTPGSGSELRSASFSAARITPLESTSEMAVTFDKYVYVNGDPVATGQFRCVPGAY	60
DB	17	LGPTPGSGSELRSASFSAARITPLESTSEMAVTFDKYVYVNGDPVATGQFRCVPGAY	76
QY	61	FFSFTAGKAPHSLSVWLRNDEVQALAFDQRRRGARRASQSMALQDVGDTWLR	120
DB	77	FFSFTAGKAPHSLSVWLRNDEVQALAFDQRRRGARRASQSMALQDVGDTWLR	136
QY	121	HGAPYALGAPGATFSGLYLVAD	143
DB	137	HGAPYALGAPGATFSGLYLVAD	159
RESULT 2			
Q8R066			

```

ID 08R066 PRELIMINARY; PRT; 326 AA.
AC 08R066;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to C1q and tumor necrosis factor related protein 4.
GN 0710001E10R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027315; AAH27315.1;
DR MGD; MGI:1914695; 0710001E10R1K.
DR InterPro; IPR001073; C1q.
DR Pfam; PF00386; C1q; 2.
DR SMART; SM00110; C1Q; 2.
DR PROSITE; PS01113; C1Q; 2.
SQ SEQUENCE 326 AA; 35057 MW; 72339172B7B1051A CRC64;

Query Match 94.8%; Score 706.5; DB 11; Length 326;
Best Local Similarity 96.5%; Pred. No. 2e-65;
Matches 138; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 LGPTPGSSSLRASFSAARTTPLEGISEMAVTEDKRVYVNIIGDFDVATGCFRCRVPCAY 60
DB 17 LGPA-GPGSSSLRASFSAARTTPLEGISEMAVTEDKRVYVNIIGDFDVATGCFRCRVPCAY 75
QY 61 FFSTFAGAPKHSLSVMLVRNDEVQALAFDEORPGARRAASOSAMLOLDYDPTVRL 120
DB 76 FFSTFAGAPKHSLSVMLVRNDEVQALAFDEORPGARRAASOSAMLOLDYDPTVRL 135
QY 121 HGAPHYALGAPGATFSGYLVYAD 143
DB 136 HGAPHYALGAPGATFSGYLVYAD 158

RESULT 3
Q9D0W2 PRELIMINARY; PRT; 205 AA.
AC 09D0W2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE 0710001E10R1K protein.
GN 0710001E10R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-C57BL/6J; TISSUE=Embryo;
RA MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirini L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.T., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guerninich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,

```

```

RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohseuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK004340; BAB23268.1;
DR MGD; MGI:1914695; 0710001E10R1K.
DR InterPro; IPR001073; C1q.
DR Pfam; PF00386; C1q; 1.
DR PRINTS; PR00007; COMPLEMENTC1Q.
DR SMART; SM00110; C1Q; 1.
DR PROSITE; PS01113; C1Q; 1.
SQ SEQUENCE 205 AA; 22190 MW; B9B237793C54786D CRC64;

Query Match 47.7%; Score 355; DB 11; Length 205;
Best Local Similarity 53.3%; Pred. No. 5.3e-29;
Matches 81; Conservative 12; Mismatches 45; Indels 14; Gaps 4;

QY 3 PTPGSSSLRASFSAARTTPLEGTS-----EMAVTFDKRVYVNIIGDFDVATGCFRCRV 56
DB 43 PARGPAPEPRSAFSAARTSLVGSDAAPGRRHPLAFDELVNIIGDFDPAAGVRCRL 102
QY 57 PGAYFFSTFGKAPKHSLSVMLVRNDEVQALAFDE--ORPGARRAASOSAMLOLDYD 114
DB 103 PGAYFFSTFGKAPKHSLSVMLVRNDEVQALAFDE--ORPGARRAASOSAMLOLDYD 159
QY 115 TWMLRLH---GAPHYALGAPGATFSGYLVYAD 143
DB 160 AVMLSHDHQGYGAYSMKXITFSGYLVYAD 191

RESULT 4
Q9DCB6 PRELIMINARY; PRT; 205 AA.
AC 09DCB6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE 0710001E10R1K protein.
GN 0710001E10R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Brain;
RA MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirini L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.T., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guerninich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohseuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK002948; BAB22473.1;
DR MGD; MGI:1914695; 0710001E10R1K.
DR InterPro; IPR001073; C1q.
DR Pfam; PF00386; C1q; 1.
DR PRINTS; PR00007; COMPLEMENTC1Q.
DR SMART; SM00110; C1Q; 1.

```

DR PROSITE: PS01113; ClQ: 1.  
SQ SEQUENCE 205 AA; 22215 MW; 56AD37793C47300 CRC64;

Query Match 46.8%; Score 349; DB 11; Length 205;  
Best Local Similarity 52.6%; Pred. No. 2.2e-28;  
Matches 80; Conservative 12; Mismatches 46; Indels 14; Gaps 4;

QY 3 PTPGSELSAFSAATTPLEGT-----ENAVTDKYVYVINGDFVATGFCRCV 56  
DB 43 PARPAPAEPRSAFSAATRLVSGDAPGPRHRLADTELVINIGDFDAAGVFCRL 102  
QY 57 PGAYFFSTAGKAPKHSVWLVNRDEVOALAPDE--QRRPGARRASOSAMQLDYGD 114  
DB 103 PGAYFFSTAGKAPKHSVWLVNRDEVOALAPDE--QRRPGARRASOSAMQLDYGD 114  
QY 115 TWMLRLH---GAPHYALGAPGATSGYLIVYAD 143  
DB 160 AVMLSHDHDGYGAVSNHGKITTSGYLIVYAD 191

## RESULT 5

ID Q9H667 PRELIMINARY; PRT; 158 AA.

AC Q9H667;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)  
DE Hypothetical protein FLJ22569.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Small intestine;  
RA Matanabe K., Kunagai A., Itakura S., Yamazaki M., Tachiro H., Ota T.,  
RA Nakamura Y., Isogai T., Sugano S.;  
RT "NEDO human cDNA sequencing project."  
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Uterus;  
RA Straube R.;  
RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AK026222; BAB15398.1; -  
DR EMBL; BC007520; AA07520.1; -  
DR InterPro: IPR001073; ClQ.  
DR Pfam; PF00386; ClQ; 1.  
DR PRINTS; PR00007; COMPLEMENTC1Q.  
DR SMART; SM00110; ClQ; 1.  
DR PROSITE; PS01113; ClQ; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 158 AA; 17625 MW; 47DB10EDD6DC9760 CRC64;

Query Match 29.8%; Score 222; DB 4; Length 158;  
Best Local Similarity 41.7%; Pred. No. 2.9e-15;  
Matches 58; Conservative 18; Mismatches 59; Indels 4; Gaps 4;

QY 7 PGSELSAFSAATTPLEGTSEMAVTFDKYVYVINGDFVATGFCRCVPGAYFFSFT 65  
DB 22 PLPQOMVAFSAARTSLAPGTLDPVFDLLNLTGETFNLQGRFNCVNGTYVFIH 81  
QY 66 AGK-APKSLSVMLVNRDEVOALAPDEQRRPGARRASOSAMQLDYGDVWLRHGA 124  
DB 82 MLKLAIVVPLVYVLMKN-EVLVSAVANDGAPD-HEHSAHAIILQFGDQIWLRLHGA 139  
QY 125 HYALGAPGATSGYLIVYAD 143  
DB 140 IYGSWMKYSTFGYLIVYOD 158

## RESULT 6

Q8TE71  
ID Q8TE71 PRELIMINARY; PRT; 1077 AA.

AC Q8TE71;  
DT 01-JUN-2002 (Tremblrel. 21, Created)  
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE EBI1.  
GN EBI1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Aerbajina W., Miller J.L.;  
RL Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AY074490; AAL71549.1; -  
DR InterPro: IPR001073; ClQ.  
DR Pfam; PF00386; ClQ; 1.  
DR SMART; SM00110; ClQ; 1.  
DR PROSITE; PS01113; ClQ; 1.  
SQ SEQUENCE 1077 AA; 120974 MW; 2B88BF3C47D032D6 CRC64;

Query Match 29.8%; Score 222; DB 4; Length 1077;  
Best Local Similarity 41.7%; Pred. No. 3.4e-14;  
Matches 58; Conservative 18; Mismatches 59; Indels 4; Gaps 4;

QY 7 PGSELSAFSAARTPL-EGTSEMAVTFDKYVYVINGDFVATGFCRCVPGAYFFSFT 65  
DB 941 PLPQOMVAFSAARTSLAPGTLDPVFDLLNLTGETFNLQGRFNCVNGTYVFIH 1000  
QY 66 AGK-APKSLSVMLVNRDEVOALAPDEQRRPGARRASOSAMQLDYGDVWLRHGA 124  
DB 1001 MLKLAIVVPLVYVLMKN-EVLVSAVANDGAPD-HEHSAHAIILQFGDQIWLRLHGA 1058  
QY 125 HYALGAPGATSGYLIVYAD 143  
DB 1059 IYGSWMKYSTFGYLIVYOD 1077

## RESULT 7

ID Q8K110 PRELIMINARY; PRT; 158 AA.

AC Q8K110;  
DT 01-OCT-2002 (Tremblrel. 22, Created)  
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Similar to hypothetical protein FLJ22569.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Uterus;  
RA Straube R.;  
RL Submitted (APR-2002) to the EMBL/Genbank/DBJ databases.  
DR EMBL; BC027523; AA07523.1; -  
DR InterPro: IPR001073; ClQ.  
DR Pfam; PF00386; ClQ; 1.  
DR PRINTS; PR00007; COMPLEMENTC1Q.  
DR SMART; SM00110; ClQ; 1.  
DR PROSITE; PS01113; ClQ; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 158 AA; 17533 MW; 86B9321C99225FCB CRC64;

Query Match 29.0%; Score 216; DB 11; Length 158;  
Best Local Similarity 41.0%; Pred. No. 1.2e-14;  
Matches 57; Conservative 19; Mismatches 59; Indels 4; Gaps 4;

QY 7 PGSELSAFSAARTPL-EGTSEMAVTFDKYVYVINGDFVATGFCRCVPGAYFFSFT 65  
DB 22 PLPQOMVAFSAARTSLAPGTLDPVFDLLNLTGETFNLQGRFNCVNGTYVFIH 81

QY 66 AGK-APRHSLSVLMVRNDEVOALAFDEQRRPGARRAASQASAMQLQDYGDTWMLRHGAP 124  
 DB 82 MLKLAIVNPLVYVIMLKN-EVLVSAYANDGAPD-HEFASNAVILQLQGDQIMVLRHGA 139  
 QY 125 HYALGAPGATFSGYLIVAD 143  
 DB 140 IYGSWMKYSTFSGYLLYOD 158

## RESULT 8

Q95UD7 PRELIMINARY; PRT; 294 AA.  
 ID Q95UD7  
 AC Q95UD7  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE 1810033K05RIK protein (RIKEN CDNA 1810033K05 gene).  
 GN 1810033K05RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_Taxid=10090;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RC STRAIN=C57BL/6J; TISSUE=Pancreas;  
 RN SEQUENCE FROM N.A.

RA Kawai T., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi S., Fukuda S.,  
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,  
 Salto T., Okazaki Y., Gogjohori T., Bono H., Kasukawa T., Salto R.,  
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 Fletschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 Blake K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 Gustincich S., Hall D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 Saeki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,  
 Suzuki H., Toyooka K., Wang K.H., Weitz C., Whiteaker C., Wilming L.,  
 Wyshak-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsunki S.,  
 Hayashizaki Y.;  
 RA "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 RN SEQUENCE FROM N.A.

RA Strauberg R.;  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK007683; BAB25187.1; -;  
 DR EMBL: BC030324; AAH30324.1; -;  
 DR MGD: MGI:191643; 1810033K05RIK.  
 DR InterPro: IPR001073; Clq.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF00386; Clq; 1.  
 DR Pfam: PF01391; Collagen; 2.  
 DR PRINTS: PR00007; COMPLEMENTC1Q.  
 DR PROSITE: PS01113; Clq; 1.  
 SQ SEQUENCE 294 AA; 30865 MW; 6D3905AE7C19E6FA CRC64;

Query Match 26.8%; Score 200; DB 11; Length 294;  
 Best Local Similarity 34.9%; Pred. No. 1,3e-12;  
 Matches 53; Conservative 30; Mismatches 45; Indels 24; Gaps 8;

QY 5 PGP---GSSELSRASFSA--RTTPEGTSEMAVTEPKYVNIIGDFVATGQRCRVPGA 59  
 DB 147 PGPSCGSSRAKSAFSAVATKSYPR--RLPIKFDKILMNGGHYNASGKFCVSPGI 203  
 QY 60 YFFSFTAGAPRHSLSVLMVRNDEVOALAFDEQRRPGARRAASQASAMQLQDYGDTWLR 119  
 DB 204 YFTFDITLA-NKHLAIGLVN-GQYRITFD--ANTGNHDAVSGSTIALKEGDEVLQ 259

QY 120 LHQA-----PHYALGAPGATFSGYLIVAD 143  
 DB 260 IFVSEQNGLFYDPYWT-----DSLFTGFLIVAD 287

## RESULT 9

Q95UD7 PRELIMINARY; PRT; 243 AA.  
 ID Q95UD7  
 AC Q95UD7  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Adiponectin.  
 GN APML.  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Macaca.  
 NCBI\_Taxid=9544;  
 RX MEDLINE=2123234; PubMed=11334417;  
 RC TISSUE=Adipose tissue;  
 RN SEQUENCE FROM N.A.

RA Hansen B.C., Matsuzawa Y.;  
 RA "Circulating concentrations of the adipocyte protein adiponectin are  
 decreased in parallel with reduced insulin sensitivity during the  
 progression to type 2 diabetes in rhesus monkeys."  
 RL Diabetes 50:1126-1133(2001).  
 DR EMBL: AF404407; AA92202.1; -;  
 DR InterPro: IPR001073; Clq.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF00386; Clq; 1.  
 DR Pfam: PF01391; Collagen; 1.  
 DR PRINTS: PR00007; COMPLEMENTC1Q.  
 DR PROSITE: PS01110; Clq; 1.  
 DR PROSITE: PS01113; Clq; 1.  
 SQ SEQUENCE 243 AA; 26264 MW; 49A45DAF2B4613FD CRC64;

Query Match 25.8%; Score 192; DB 6; Length 243;  
 Best Local Similarity 31.8%; Pred. No. 6.8e-12;  
 Matches 47; Conservative 29; Mismatches 54; Indels 18; Gaps 6;

QY 5 PGPSSSELSRASFSAARTTPEGTSEMAVTEPKYVNIIGDFVATGQRCRVPGAFESF 64  
 DB 103 PGEAAYVRSASFVGLTYVT-VNMPPIRFKIRYNOQNHVDSGTFKHCNIPELYTFAY 161  
 QY 65 TAGRAPH-----KSLSVLMVRNDEVOALAFDEQRRPGARRAASQASAMQLQDYGDTWLR 119  
 DB 162 -----HITVYMKQVKVSLFK-KDKAMLFTYDQGENNVNQ-ASGSVILHLHEVDQVWLQ 213  
 QY 120 LHGAH-----YALGAPGATFSGYLIVAD 143  
 DB 214 VYGEGERNGLYADNDNDSTFTGFLIVAD 241

## RESULT 10

Q95UD7 PRELIMINARY; PRT; 243 AA.

ID Q95UD7  
 AC Q95UD7  
 DT 01-OCT-2002 (TREMBLrel. 22, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Clq and tumor necrosis factor related protein 5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_Taxid=9606;  
 RN SEQUENCE FROM N.A.



RC TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (May-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC029485; AAH29485.1; -  
DR InterPro; IPR001073; Clq.  
DR InterPro; IPR000087; Collagen.  
DR Pfam; PF00386; Clq; 1.  
DR Pfam; PF01391; Collagen; 1.  
DR PRINTS; PRO0007; COMPLEMENT1Q.  
DR SMART; SM00110; Clq; 1.  
SQ SEQUENCE 243 AA; 25326 MW; 6D9306A0EB21B44A CRC64;

Query Match 25.4%; Score 189; DB 4; Length 243;  
Best Local Similarity 35.3%; Pred. No. 1.4e-11;  
Matches 53; Conservative 28; Mismatches 57; Indels 12; Gaps 6;

QY 2 GPTGPGSSSL--SSAFSAART--TPLEGTSMAVTFDQVYVNIIGDPPVATGQRCRVPG 58  
DB 90 GPTGPGGCSVPSPSAFAKSESSESVPPSDAPLPFDVLYVNEQCHYAVTGKTCQVPG 149  
QY 59 AYFESFTGAKAPHSLSVMLVNRNDEVQAL--AFDEQRPGARRAASOSAMLOLDYGDPTV 116  
DB 150 VYFAVVAH--TVYRASLQFDLVKNGESIASFFQFGWPKPA---SLSGAMVRLPEPDQV 205  
QY 117 WRLHGAHP--YALGAPGATFSGLYVAD 143  
DB 206 WVQGVGDYIGIYASIKTSTFSGFLVYSD 235

RESULT 11  
Q8BRW2 PRELIMINARY; PRT; 247 AA.

AC Q8BRW2;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Adipocyte complement related protein of 30 kDa.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573 (2002).  
RL EMBL; AK041214; BAC30866.1; -  
SQ SEQUENCE 247 AA; 26751 MW; 0D3FA64C789CAEF3 CRC64;

Query Match 24.8%; Score 185; DB 11; Length 247;  
Best Local Similarity 31.8%; Pred. No. 3.7e-11;

Matches 47; Conservative 29; Mismatches 54; Indels 18; Gaps 7;

QY 5 PGPGSSSLRSAPSAARTTPLEGTSMAVTFDQVYVNIIGDPPVATGQRCRVPGAPSPF 64  
DB 107 PGEAAVYKRSAPFSGLETRV--VENVPIRFTKIFLNQNNHGYDGTGKTYCNIPGLYYSY 165  
QY 65 TAGKAPH-----KSLSVMLVNRNDEVQALAFDEQRPGARRAASOSAMLOLDYGDPTWLR 119  
DB 166 -----HTVYMKDVKSLFK-KDAVAFYFTDQYGEKAVDQ--ASSSVLHLHEVGQVWLQ 217  
QY 120 LHG-APH--YALGAPGATFSGLYVAD 143  
DB 218 VYGDGDNHGLYADNVNDSFTFGFLYHD 245

RESULT 12  
Q8R002 PRELIMINARY; PRT; 243 AA.

AC Q8R002;  
DT 01-JUN-2002 (TREMBLrel. 21, Created)  
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Similar to DKFZP586B0621 protein (Hypothetical 25.4 kDa protein).  
GN Clotnfs.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.

RA Strausberg R.;  
RL Submitted (Feb-2002) to the EMBL/GenBank/DBJ databases.  
RN 121  
RP SEQUENCE FROM N.A.

RA Strausberg R.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC023068; AAH23068.1; -  
DR EMBL; BC025174; AAH25174.1; -  
DR MGI; MGI:2385958; Clotnfs.  
DR InterPro; IPR001073; Clq.  
DR InterPro; IPR000087; Collagen.  
DR Pfam; PF00386; Clq; 1.  
DR Pfam; PF01391; Collagen; 1.  
DR SMART; SM00110; Clq; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 243 AA; 25420 MW; 498129CD051DB97B CRC64;

Query Match 24.4%; Score 181.5; DB 11; Length 243;  
Best Local Similarity 37.6%; Pred. No. 8.4e-11;  
Matches 56; Conservative 27; Mismatches 55; Indels 11; Gaps 6;

QY 1 LGPTGPGSSSLRSAPSAART--TPLEGTSMAVTFDQVYVNIIGDPPVATGQRCRVGA 59  
DB 92 IGPA-GGCSVPSPSAFAKSESSESVPPSDAPLPFDVLYVNEQCHYAVTGKTCQVPG 150  
QY 60 YFESFTGAKAPHSLSVMLVNRNDEVQALAFDEQRPGARRAASOS--AMLOLDYGDPTW 117  
DB 151 YVFAVVAH--TVYRASLQFDLVKNG--QSISFFQYFGWPKPASP.LSGAMVRLPEPDQV 206  
QY 118 WRLHGAHP--YALGAPGATFSGLYVAD 143  
DB 207 WVQGVGDYIGIYASIKTSTFSGFLVYSD 235

RESULT 13  
Q8BKRO PRELIMINARY; PRT; 264 AA.

AC Q8BKRO;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Weekly similar to complement-C1Q tumor necrosis factor-related  
DE protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Body;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573 (2002).  
RL EMBL; AK051058; BAC34512.1; -  
SQ SEQUENCE 264 AA; 29062 MW; 054B50E7D2E56826 CRC64;

Query Match 24.2%; Score 180.5; DB 11; Length 264;  
Best Local Similarity 35.7%; Pred. No. 1.2e-10;

Matches 56; Conservative 15; Mismatches 61; Indels 25; Gaps 6;

QY 2 GPTPGGSS--ELRSNFAARTTPEGTSE-NAVTFDKVYVNIIGDPDVATGQFRCRVG 58  
 DB 116 GGAGSGSSCCQTHYAFSVGRKTGLHSSNFSLIFDRFVNTDGHFDMAITGSFAPLRG 175  
 QY 59 AYFSEFTACKAPHKSLVMLVNRDEVOALAFDEORRRPGARR-AASQASMLQLDYDGTW 117  
 DB 176 LTFEFSINVHSMYKETYVHAIHNEQAVVIL---YAQPSERSIMOSQSVMLPLVPGDRVW 231  
 QY 118 LRL-----HGAPHYALGAGATFSGVLYVAD 143  
 DB 232 VALFKRENGIYSDVDVTY-----ITFSGHLIRAE 262

RESULT 14

Q8K3R4 PRELIMINARY; PRT; 244 AA.

AC 08K3R4; 01-OCT-2002 (Tremblrel. 22, Created)  
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE 30 kDa adipocyte complement-related protein.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RA Berg A.H., Scherer P.B.;  
 RT "Cloning of the rat homolog to murine adipocyte complement-related  
 RT protein of 30 kDa."  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY033885; AAK61608.1; -  
 DR InterPro; IPR001073; Clq.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; Clq; 1.  
 DR Pfam; PF01391; Collagen; 1.  
 DR PRINTS; PR00007; COMPLEMENTC1Q.  
 DR ProDom; PD000007; Collagen; 1.  
 DR SMART; SM00110; C1Q; 1.  
 DR PROSITE; PS01113; C1Q; 1.  
 KW Collagen.  
 SQ SEQUENCE 244 AA; 26410 MW; 75B2B1DF68E2633E CRC64;

Query Match 24.2%; Score 180; DB 11; Length 244;  
 Best Local Similarity 30.4%; Pred. No. 1.2e-10;  
 Matches 45; Conservative 29; Mismatches 56; Indels 18; Gaps 6;

QY 5 PGPSSSELRSAPFAARTTPEGTSEMAVTFDKVYVNIIGDPDVATGQFRCRVGAYFFSF 64  
 DB 104 PGEAAVYRSAPFSGLETRV-VPNPPIRFKITYNQNHDSGTGKFCYNIPGLYFYSY 162  
 QY 65 TAGKAPH-----KSLSYMLVNRDEVOALAFDEORRRPGARRAASQASMLQLDYDGTWLR 119  
 DB 163 -----HITVYMKDVKVSLEFK-KDKAVLFTYDQYQEKKNVDQ-ASGSVLLHLLEVGDQVWLQ 214  
 QY 120 LHGAPH---YALGAPGATFSGVLYVAD 143  
 DB 215 VTGEGDNGLYADNVNDSTFTGFLYHD 242

RESULT 15

Q95MO4 PRELIMINARY; PRT; 240 AA.

AC 095MO4; 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Adipose tissue-specific protein adipo Q.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 NCBI\_TaxId=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2136933; PubMed=11382781;  
 RA Sato C., Yasukawa Z., Honda N., Matsuda T., Kitajima K.;  
 RT "Identification and Adipocyte Differentiation-dependent Expression of  
 RT the Unique Distal Acid Residue in an Adipose Tissue-specific  
 RT Glycoprotein, Adipo Q."  
 RT J. Biol. Chem. 276:28849-28856 (2001).  
 DR EMBL; AF269230; AAK58902.1; -  
 DR InterPro; IPR001073; Clq.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; Clq; 1.  
 DR Pfam; PF01391; Collagen; 1.  
 DR PRINTS; PR00007; COMPLEMENTC1Q.  
 DR ProDom; PD000007; Collagen; 1.  
 DR SMART; SM00110; C1Q; 1.  
 DR PROSITE; PS01113; C1Q; 1.  
 KW Collagen.  
 SQ SEQUENCE 240 AA; 26091 MW; C6253BA803B9A668 CRC64;

Query Match 24.0%; Score 179; DB 6; Length 240;  
 Best Local Similarity 31.5%; Pred. No. 1.5e-10;  
 Matches 46; Conservative 29; Mismatches 53; Indels 18; Gaps 7;

QY 5 PGPSSSELRSAPFAARTTPEGTSEMAVTFDKVYVNIIGDPDVATGQFRCRVGAYFFSF 64  
 DB 99 PGEAAVYRSAPFSGLETRV-VPNPPIRFKITYNQNHDSGTGKFCYNIPGLYFYSY 157  
 QY 65 TAGKAPH-----KSLSYMLVNRDEVOALAFDEORRRPGARRAASQASMLQLDYDGTWLR 119  
 DB 158 -----HITVYMKDVKVSLEFK-KDKAVLFTYDQYQEKKNVDQ-ASGSVLLHLLEVGDQVWLQ 209  
 QY 120 LH-GAPH---YALGAPGATFSGVLYV 141  
 DB 210 VTGEGDNGLYADNVNDSTFTGFLY 235

Search completed: January 12, 2004, 08:19:09  
 Job time : 20.5946 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:09 ; Search time 21.5499 Seconds

(without alignments)  
1053.272 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_17\_159

Perfect score: 745  
Sequence: 1 LGPTPGSGSELRSAARSAR.....PHYALCAPATRSGLVYAD 143

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 segs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: A\_Geneseq\_19Jun03:\*  
2: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
3: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
4: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
5: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
6: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
7: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
8: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
9: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
10: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
11: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
12: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
13: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
14: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
15: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
16: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
17: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
18: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
19: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
20: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
21: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
22: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
23: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
24: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
25: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	745	100.0	299	23	ABG70385	Adipocyte compleme
2	745	100.0	329	22	ABG70385	Human ZACRP4. Hom
3	737	98.9	284	23	ABG70384	Adipocyte compleme
4	737	98.9	348	22	ABG70384	Human MANGO 245 pr
5	733	98.4	334	22	ABG70383	Human MANGO 245 ma
6	726	97.4	329	22	ABG70383	Human novel secret
7	708	95.0	329	22	ABG70383	Monkey MANGO 245 p
8	707.5	95.0	199	22	ABG70383	Murine MANGO 245 p
9	646	86.7	126	22	ABG70383	Human MANGO 245 C1

10	639	85.8	126	22	ABG70385	Murine MANGO 245 C
11	635	85.2	126	22	ABG70385	Human MANGO 245 ex
12	632	83.5	126	22	ABG70385	Monkey MANGO 245 C
13	618	83.0	123	21	ABG70385	Human ORFY ORF1953
14	448	60.1	221	23	ABG70383	Adipocyte compleme
15	426.5	57.2	130	22	ABG70386	Mature monkey MANG
16	341.5	45.8	133	23	ABG70386	Adipocyte compleme
17	320	43.0	192	22	ABG70386	Human MANGO 245 cy
18	314	42.1	134	22	ABG70386	Monkey MANGO 245 C
19	305	40.9	127	22	ABG70386	Human secreted pro
20	301.5	40.5	117	22	ABG70386	Human MANGO 245 C1
21	222	29.8	202	22	ABG70386	Human polypeptide
22	222	29.8	709	23	ABG70386	Human pancreatic c
23	222	29.8	710	23	ABG70386	Human genest metab
24	222	29.8	746	20	ABG70386	Human lung tumour
25	222	29.8	746	21	ABG70386	Human lung tumour
26	222	29.8	746	22	ABG70386	Human lung tumour
27	222	29.8	800	22	ABG70386	Human lung tumour
28	222	29.8	908	22	ABG70386	Human protein SEQ
29	222	29.8	909	22	ABG70386	Novel human protei
30	222	29.8	957	22	ABG70386	Novel human protei
31	222	29.8	958	22	ABG70386	Novel human protei
32	222	29.8	992	22	ABG70386	Novel human protei
33	222	29.8	993	22	ABG70386	Novel human protei
34	222	29.8	1042	22	ABG70386	Novel human protei
35	222	29.8	1043	22	ABG70386	Novel human protei
36	222	29.8	1091	22	ABG70386	Novel human protei
37	222	29.8	1092	22	ABG70386	Novel human protei
38	222	29.8	1126	22	ABG70386	Novel human protei
39	222	29.8	1127	22	ABG70386	Novel human protei
40	209	28.1	710	22	ABG70386	Human protein SEQ
41	200	26.8	285	21	ABG70386	Murine ACP30RIM.
42	200	26.8	294	21	ABG70386	Rat skin cell prot
43	200	26.8	294	22	ABG70386	Skin cell protein.
44	200	26.8	294	23	ABG70386	Rat protein isolat
45	197	26.4	285	21	ABG70386	Rat skin cell secr

## ALIGNMENTS

RESULT 1	ABG70385	standard; Protein; 299 AA.
ID	ABG70385	
XX	ABG70385	
AC	ABG70385	
DT	05-NOV-2002	(first entry)
XX	ADIPocyte complement-related C1q Tumour Necrosis Factor-like protein #3.	
DE	ADIPocyte complement-related C1q Tumour Necrosis Factor-like protein #3.	
KW	Human; NOVX; NOVX-associated disorder; cardiomyopathy; atherosclerosis;	
KW	cell signal processing; metabolic pathway modulation; metabolic disorder;	
KW	obesity; diabetes; infectious disease; neurodegenerative disorder; acne;	
KW	Alzheimer's disease; Parkinson's disease; immune disorder; cancer;	
KW	haematopoietic disorder; cirrhosis; pancreatitis; learning defect;	
KW	memory defect; infertility; congenital heart defect; hair growth;	
KW	pigmentation disorder; endocrine disorder; respiratory disease; health;	
KW	gastro-intestinal disease; reproductive; neurological disease;	
KW	bone marrow transplantation; endocrine disease; allergy; inflammation;	
KW	neurological disorder; urinary system disorder; age-related disorder;	
KW	neuropsychiatric disorder; EGF-related protein; SCUBE1; TEN-14;	
KW	adipocyte complement-related C1q tumour necrosis factor; out at first;	
KW	beta adrenergic receptor kinase; EphA6/ehk-2; glucose transporter;	
KW	type 1a membrane sushi-containing domain; butyrophilin;	
KW	type 1a membrane sushi-containing domain; butyrophilin;	
OS	Homo sapiens.	
PN	WO200257453-A2.	
XX	25-JUL-2002.	
PD		
XX		

PF 19-DEC-2001; 2001WO-US50331.  
XX  
XX 19-DEC-2000; 2000US-265704P.  
PR 20-DEC-2000; 2000US-257319P.  
PR 02-MAY-2001; 2001US-286153P.  
PR 29-MAY-2001; 2001US-294075P.  
PR 24-JUL-2001; 2001US-307506P.  
PR 10-AUG-2001; 2001US-311590P.  
PR 10-AUG-2001; 2001US-311613P.  
PR 29-AUG-2001; 2001US-315617P.  
PR 14-SEP-2001; 2001US-322358P.  
XX  
XX (CURA-) CURAGEN CORP.  
PI Gangoli BA, Patturajan M, Vernet CAM, Malvankar UM, Kekuda R;  
PI Stone DJ, Anderson D, Shimkets RA, Burgess CE, Zernhusen BD, Liu X;  
PI Spytek KA, Casman SJ, Boldog FU, Smithson G, Li L, Ji W;  
XX  
XX WPI; 2002-590744/63.  
DR N-PSDB; ABSS2097.  
XX  
XX  
PS Claim 1; Page 31; 318pp; English.  
XX  
XX The present invention relates to new NOVX polypeptides. The invention is  
CC useful for treating or preventing a NOVX-associated disorder such as  
CC cardiomyopathy or atherosclerosis, where the disorder is related to cell  
CC signal processing and metabolic pathway modulation in a subject.  
CC preferably human. The invention is also useful for treating metabolic  
CC disorders (e.g. Alzheimer's disease, infectious disease, neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), immune  
CC disorders, haematopoietic disorders and various cancers. The molecules of  
CC the invention are also useful for treating or preventing cirrhosis,  
CC pancreatitis, learning and memory defects, infertility, congenital heart  
CC defects, acne, hair growth, pigmentation disorders, endocrine disorders,  
CC respiratory disease, gastro-intestinal diseases, reproductive health,  
CC neurological diseases, bone marrow transplantation, endocrine diseases,  
CC allergy and inflammation, nephrological disorders, urinary system  
CC disorders, neuropsychiatric disorders and age-related disorders.  
CC The present amino acid sequence represents a NOVX protein of the  
CC invention.  
XX  
XX  
SQ Sequence 299 AA;  
Query Match 100.0%; Score 745; DB 23; Length 299;  
Best Local Similarity 100.0%; Pred. No. 1.5e-76;  
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LGPTPGSSSELRSASFAARTTPLEGTSEMAVTFDKYVNIIGDFDVAATGQFRCPVGA 60  
DB 17 LGPTPGSSSELRSASFAARTTPLEGTSEMAVTFDKYVNIIGDFDVAATGQFRCPVGA 76  
QY 61 FFSTTAGKAPKHSISVNLVNRDEVQALAFDEORRPGARRAASOSAMLOLDYGTWRL 120  
DB 77 FFSTTAGKAPKHSISVNLVNRDEVQALAFDEORRPGARRAASOSAMLOLDYGTWRL 136  
QY 121 HGAPHYALGAPGATFSGLVYAD 143  
DB 137 HGAPHYALGAPGATFSGLVYAD 159

DE Human ZACRP4.  
XX  
XX Human; zacrpa; complement factor C1q domain; chromosome 11q11;  
XX energy balance; cellular metabolic reaction; autocrine factor;  
XX development; cell proliferation; differentiation; cell survival.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH Peptide /label= Signal\_peptide  
FT /label= C1q\_domain\_#1  
FT /label= C1q\_domain\_#1  
FT /label= C1q\_domain\_#2  
FT Domain  
PN W0200102565-A2.  
XX  
XX 11-JAN-2001.  
XX  
XX 28-JUN-2000; 2000WO-US17692.  
XX  
XX  
XX 01-JUL-1999; 99US-0346502.  
XX  
XX (ZYMO) ZYMOGENETICS INC.  
XX  
XX Holloway JL, Lok S;  
PI  
XX  
XX WPI; 2001-138140/14.  
DR N-PSDB; AAF28672.  
XX  
XX Novel secreted protein ZACRP4 polypeptides having tandem C1q globular  
PT domains, useful for studying cell-cell communication and regulation of  
PT cellular processes -  
XX  
XX  
PS Claim 1; Page 77-78; 82pp; English.  
XX  
XX The present sequence is human ZACRP4 protein. ZACRP4 protein has two  
CC complement factor C1q domains. The ZACRP4 gene is located on human  
CC chromosome 11q11. The ZACRP4 coding sequence and protein have a number of  
CC uses described in the specification, including, modulation of energy  
CC balance and cellular metabolic reactions in mammals. In addition, ZACRP4  
CC protein is useful as an autocrine factor, particularly during  
CC development, in mediating the processes of an organism, in regulating  
CC cellular processes such as cell proliferation and/or differentiation,  
CC cell survival and energy balance.  
XX  
XX  
SQ Sequence 329 AA;  
Query Match 100.0%; Score 745; DB 22; Length 329;  
Best Local Similarity 100.0%; Pred. No. 1.7e-76;  
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LGPTPGSSSELRSASFAARTTPLEGTSEMAVTFDKYVNIIGDFDVAATGQFRCPVGA 60  
DB 17 LGPTPGSSSELRSASFAARTTPLEGTSEMAVTFDKYVNIIGDFDVAATGQFRCPVGA 76  
QY 61 FFSTTAGKAPKHSISVNLVNRDEVQALAFDEORRPGARRAASOSAMLOLDYGTWRL 120  
DB 77 FFSTTAGKAPKHSISVNLVNRDEVQALAFDEORRPGARRAASOSAMLOLDYGTWRL 136  
QY 121 HGAPHYALGAPGATFSGLVYAD 143  
DB 137 HGAPHYALGAPGATFSGLVYAD 159

RESULT 2  
AAB61606  
ID AAB61606 standard; Protein; 329 AA.  
XX  
XX AAB61606;  
XX  
XX 05-APR-2001 (first entry)  
XX  
XX

RESULT 3  
ABG70384  
ID ABG70384 standard; Protein; 284 AA.  
XX  
XX ABG70384;  
XX  
XX 05-NOV-2002 (first entry)  
XX  
XX

XX Adipocyte complement-related C1q Tumour Necrosis Factor-like protein #2.  
DE  
XX  
XX Human, NOXV; NOXV-associated disorder; cardiomyopathy; atherosclerosis;  
KM cell signal processing; metabolic pathway modulation; atherosclerosis;  
KM obesity; diabetes; infectious disease; neurodegenerative disorder; acne;  
KM Alzheimer's disease; Parkinson's disease; immune disorder; cancer;  
KM hematopoietic disorder; cirrhosis; pancreatitis; learning defect;  
KM memory defect; infertility; congenital heart defect; hair growth;  
KM pigmentation disorder; endocrine disorder; respiratory disease; health;  
KM gastro-intestinal disease; reproductive; neurological disease;  
KM bone marrow transplantation; endocrine disease; allergy; inflammation;  
KM neurological disorder; urinary system disorder; age-related disorder;  
KM neuropsychiatric disorder; EGF-related protein; SCUB1; TEN-M4;  
KM adipocyte complement-related C1q tumour necrosis factor; out at first;  
KM beta adrenergic receptor kinase; EphA6/ehk-2; glucose transporter;  
KM type 1a membrane sushi-containing domain; butyrophilin;  
XX  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200257453-A2.  
PN  
XX  
XX 25-JUL-2002.  
PD  
XX  
XX 19-DEC-2001; 2001WO-US50331.  
PF  
XX  
XX 19-DEC-2000; 2000US-265704P.  
PR  
XX 20-DEC-2000; 2000US-257314P.  
PR 02-MAY-2001; 2001US-288153P.  
PR 29-MAY-2001; 2001US-294075P.  
PR 24-JUL-2001; 2001US-307506P.  
PR 10-AUG-2001; 2001US-311590P.  
PR 10-AUG-2001; 2001US-311613P.  
PR 29-AUG-2001; 2001US-315617P.  
PR 14-SEP-2001; 2001US-322358P.  
XX  
XX  
XX (CURA-) CURAGEN CORP.  
PA  
XX Gangoli EA, Patturajan M, Vernet CAM, Malvankar UM, Kekuda R;  
PI Stone DJ, Anderson D, Shimkets RA, Burgess CE, Zernhusen BD, Liu X;  
PI Szytek KA, Casman SJ, Boldog FL, Smithson G, Li L, Ji W;  
XX WPI; 2002-550744/63.  
DR N-PSDB; ABSS2096.  
DR  
XX  
XX Novel isolated NOXV polypeptide useful for treating cardiomyopathy,  
PT atherosclerosis, metabolic disorders, diabetes, obesity, infectious  
PT disease, anorexia, neurodegenerative disorders, Alzheimer's disease or  
PT cancer -  
XX  
XX  
XX Claim 1; Page 30; 318pp; English.  
PS  
XX  
XX The present invention relates to new NOXV polypeptides. The invention is  
CC useful for treating or preventing a NOXV-associated disorder such as  
CC cardiomyopathy or atherosclerosis, where the disorder is related to cell  
CC signal processing and metabolic pathway modulation in a subject,  
CC preferably human. The invention is also useful for treating metabolic  
CC disorders (e.g. obesity), diabetes, infectious disease, neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), immune  
CC disorders, hematopoietic disorders and various cancers. The molecules of  
CC the invention are also useful for treating or preventing cirrhosis,  
CC pancreatitis, learning and memory defects, infertility, congenital heart  
CC defects, acne, hair growth, pigmentation disorders, endocrine diseases,  
CC respiratory disease, gastro-intestinal diseases, reproductive, health,  
CC neurological diseases, bone marrow transplantation, endocrine diseases,  
CC allergy and inflammation, neurological disorders, urinary system  
CC disorders, neuropsychiatric disorders and age-related disorders.  
CC The present amino acid sequence represents a NOXV protein of the  
CC invention.  
XX  
XX Sequence 284 AA;

Query Match 98.9%; Score 737; DB 23; Length 284;  
Best Local Similarity 99.3%; Pred. No. 1.2e-75;  
Matches 142; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 LGFTPGGSSLSAFSAARTTPLEGSEMAVTFDKYVNIIGDPDVATGQFCRVPGAY 60  
Db 17 LGFTPGGSSLSAFSAARTTPLEGSEMAVTFDKYVNIIGDPDVATGQFCRVPGAY 76  
Qy 61 FFSFTAGKAPHKSLVYLVRNDEVQALAFDEQRPRGARRAASQSAWLQLDYGDYTWLRL 120  
Db 77 FFSFTAGKAPHKSLVYLVRNDEVQALAFDEQRPRGARRAASQSAWLQLDYGDYTWLRL 136  
Qy 121 HGAPYALGAPGATFSGYLVYAD 143  
Db 137 HGAPYALGAPGATFSGYLVYAD 159  
RESULT 4  
ID AAB61423 standard; protein; 348 AA.  
XX  
XX AAB61423;  
AC  
XX  
XX 04-APR-2001 (first entry)  
DT  
XX  
XX Human MANGO 245 protein.  
DE  
XX  
XX TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
KM autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
KM pancreatic; skeletal; muscle.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200100672-A1.  
PN  
XX  
XX 04-JAN-2001.  
PD  
XX  
XX 29-JUN-2000; 2000WO-US18184.  
PF  
XX  
XX 29-JUN-1999; 99US-0342687.  
PR  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
PA  
XX  
XX Holtzman DA, Barnes TM, Fraser CC, Sharp JD;  
PI WPI; 2001-050127/06.  
DR  
XX  
XX Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
PT disorders (e.g. jaundice) -  
XX  
XX  
XX Claim 1; Fig 23; 262pp; English.  
PS  
XX  
XX The present invention relates to cDNAs encoding TANGO 244,  
CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
CC The nucleic acids, proteins and protein modulators are useful for  
CC treating colonic disorders, inflammatory diseases, tumors,  
CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
CC allergic diseases, cardiovascular diseases, brain disorders,  
CC degenerative diseases placental, pancreatic, skeletal and muscle  
CC disorders.  
XX  
XX  
XX Sequence 348 AA;  
SQ  
Query Match 98.9%; Score 737; DB 22; Length 348;  
Best Local Similarity 99.3%; Pred. No. 1.5e-75;  
Matches 142; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 LGFTPGGSSLSAFSAARTTPLEGSEMAVTFDKYVNIIGDPDVATGQFCRVPGAY 60  
Db 17 LGFTPGGSSLSAFSAARTTPLEGSEMAVTFDKYVNIIGDPDVATGQFCRVPGAY 76

QY 61 FFSTAGKAPKHSISVLMVRNDEVQALAFDEQRPGARRAASQSAMQLDYGDTWRL 120  
DB 77 FFSTAGKAPKHSISVLMVRNDEVQALAFDEQRPGARRAASQSAMQLDYGDTWRL 136  
QY 121 HGAPHYALGAPGATFSGLYVAD 143  
DB 137 HGAPOYALGAPGATFSGLYVAD 159

RESULT 5  
AAB61466  
ID AAB61466 standard; protein; 334 AA.  
XX  
AC AAB61466;  
XX  
DT 04-APR-2001 (first entry)  
XX  
DE Human MANGO 245 mature protein.  
XX  
KM TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
KM autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
KM pancreatic; skeletal; muscle.  
XX  
OS Homo sapiens.  
XX  
PN WO200100672-A1.  
XX  
PD 04-JAN-2001.  
XX  
PF 29-JUN-2000; 2000WO-US18184.  
XX  
PR 29-JUN-1999; 99US-0342687.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Holzman DA, Barnes TM, Frazer CC, Sharp JD;  
XX  
DR WPI; 2001-050127/06.  
XX  
PT Isolated cDNA encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
PT disorders (e.g. jaundice) -  
XX  
XX  
PS Disclosure: Page 242; 262pp; English.  
XX  
CC The present invention relates to cDNAs encoding TANGO 244,  
CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
CC The nucleic acids, proteins and protein modulators are useful for  
CC treating colonic disorders, inflammatory diseases, tumors,  
CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
CC allergic diseases, cardiovascular diseases, brain disorders,  
CC degenerative diseases placental, pancreatic, skeletal and muscle  
CC disorders.  
XX  
SQ Sequence 334 AA;  
XX

Query Match 98.4%; Score 733; DB 22; Length 334;  
Best Local Similarity 98.6%; Pred. No. 4, 1e-75;  
Matches 141; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 LGPTGGPSSELRSASFSAKRTPLRGTSMAVTPDKVYVNIIGDPAVATGQFRCRVPAV 60  
DB 3 LGPTGGPSSELRSAGFSAKRTPLRGTSMAVTPDKVYVNIIGDPAVATGQFRCRVPAV 62  
QY 61 FFSTAGKAPKHSISVLMVRNDEVQALAFDEQRPGARRAASQSAMQLDYGDTWRL 120  
DB 63 FFSTAGKAPKHSISVLMVRNDEVQALAFDEQRPGARRAASQSAMQLDYGDTWRL 122  
QY 121 HGAPHYALGAPGATFSGLYVAD 143  
DB 123 HGAPOYALGAPGATFSGLYVAD 145

RESULT 6  
ABG79643  
ID ABG79643 standard; protein; 329 AA.  
XX  
AC ABG79643;  
XX  
DT 15-NOV-2002 (first entry)  
XX  
DE Human novel secreted protein SECP19, Incyte ID No. 931619CD1.  
XX  
KM Human; SECP; secreted protein; cell proliferative disorder;  
KM actinic keratosis; arteriosclerosis; bursitis; hepatitis; cancer;  
KM autoimmune disorder; inflammatory disorder; AIDS; asthma; allergy;  
KM acquired immunodeficiency syndrome; anaemia; atopic dermatitis;  
KM cardiovascular disorder; congestive heart failure; vascular tumour;  
KM ischaemic heart disease; myocardial infarction; epilepsy; stroke;  
KM hypertensive heart disease; neurological disorder; cerebral neoplasm;  
KM Alzheimer's disease; developmental disorder; renal tubular acidosis;  
KM Cushing's syndrome; Duchenne muscular dystrophy; hypothyroidism;  
KM Becker muscular dystrophy.  
XX  
OS Homo sapiens.  
XX  
PN WO200262841-A2.  
XX  
PD 15-AUG-2002.  
XX  
PF 28-JAN-2002; 2002WO-US02616.  
XX  
PR 02-FEB-2001; 2001US-266195P.  
PR 08-FEB-2001; 2001US-267924P.  
PR 09-FEB-2001; 2001US-267816P.  
PR 09-FEB-2001; 2001US-268112P.  
PR 26-FEB-2001; 2001US-271639P.  
PR 07-SEP-2001; 2001US-317818P.  
PR 21-DEC-2001; 2001US-343553P.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
XX  
PI Tang TY, Yue H, Gandhi AR, Yao MG, Warren BA, Ding L, Duggan BM;  
PI Xu Y, Yang U, Thangavelu K, Lal PG, Honchell CD, Walla NK, Lee S;  
PI Lee EA, Richardson TW, Baughn WR, Elliott VS;  
XX  
XX  
DR WPI; 2002-657522/70.  
DR N-PSDB; ABS64954.  
XX  
XX  
PT New human secreted proteins and nucleic acids useful in diagnosing,  
PT treating and preventing cell proliferative, autoimmune/inflammatory,  
PT cardiovascular, neurological, and developmental disorders -  
XX  
XX  
PS Claim 1; Page 140; 158pp; English.  
XX  
CC The invention relates to twenty four human secreted proteins  
CC (SECP1-24), proteins 90% identical to chem and active fragments of them.  
CC Also included are nucleic acids encoding the SECP proteins, a recombinant  
CC polynucleotide comprising a promoter sequence operably linked to the  
CC nucleic acid, a cell transformed with the recombinant polynucleotide, an  
CC a transgenic organism comprising the recombinant polynucleotide, an  
CC anti-SECP antibody, and screening for ant/agonists and modulators of  
CC SECP function or expression. The SECP proteins and nucleic acids are  
CC useful in the diagnosis, treatment and prevention of cell proliferative  
CC (e.g. actinic keratosis, arteriosclerosis, bursitis, hepatitis or  
CC cancer), autoimmune/inflammatory (e.g. AIDS (acquired immunodeficiency  
CC syndrome), asthma, anaemia, allergies or atopic dermatitis),  
CC cardiovascular (e.g. congestive heart failure, ischaemic heart disease,  
CC myocardial infarction, hypertensive heart disease, or vascular tumours),  
CC neurological (e.g. epilepsy, stroke, cerebral neoplasms, or Alzheimer's  
CC disease), and developmental (e.g. renal tubular acidosis, Cushing's  
CC syndrome, Duchenne and Becker muscular dystrophy, or hypothyroidism)  
CC disorders. Many other diseases and disorders are listed in the  
CC specification. These may also be used in assessing the effects of  
CC exogenous compounds on the expression of nucleic acid and amino acid

CC sequences of the secreted proteins. The present sequence represents a  
CC SSCP protein of the invention.  
XX  
SQ Sequence 329 AA;  
Query Match 97.4%; Score 726; DB 23; Length 329;  
Best Local Similarity 98.6%; Pred. No. 2.5e-74;  
Matches 141; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 LGPTPGSSSELSAFSAARTTPLEGTSEMAVTFDKYVNIIGDFDVATGQFRCRVGAY 60  
DB 17 LGPTPGSSSELSAFSAARTTPLEGTSEMAVTFDKYVNIIGDFDVATGQFRCRVGAY 76  
QY 61 FFSFTAGKAPHKSLSWLVNRNDEVQALAFDEQRPPGARRAASQSAMLQLDYGDPTWRL 120  
DB 77 FFSFTAGKAPHKSLSWLVNRNDEVQALAFDEQRPPGARRAASQSAMLQLDYGDPTWRL 136  
QY 121 HGAPHYALGAPGATFSGYLIVYAD 143  
DB 137 LGAPHYALGAPGATFSGYLIVYAD 159  
RESULT 7  
AAB61424  
ID AAB61424 standard; protein; 329 AA.  
AC AAB61424;  
XX  
XX 04-APR-2001 (first entry)  
DT  
DE Monkey MANGO 245 protein.  
XX  
KW TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
KM autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
XX pancreatic; skeletal; muscle.  
OS Catarrhini sp.  
XX  
PN WO200100672-A1.  
XX  
PD 04-JAN-2001.  
XX  
PF 29-JUN-2000; 2000WO-US18184.  
XX  
PR 29-JUN-1999; 99US-0342687.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;  
XX  
DR WPI; 2001-050127/06.  
XX  
PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
PT disorders (e.g. jaundice) -  
XX  
PS Claim 1; Fig 25; 262pp; English.  
XX  
CC The present invention relates to cDNAs encoding TANGO 244,  
CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
CC The nucleic acids, proteins and protein modulators are useful for  
CC treating colonic disorders, inflammatory diseases, tumors,  
CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
CC allergic diseases, cardiovascular diseases, brain disorders,  
CC degenerative diseases placental, pancreatic, skeletal and muscle  
CC disorders.  
XX  
SQ Sequence 329 AA;  
Query Match 95.0%; Score 708; DB 22; Length 329;  
Best Local Similarity 95.1%; Pred. No. 2.9e-72;  
Matches 136; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 LGPTPGSSSELSAFSAARTTPLEGTSEMAVTFDKYVNIIGDFDVATGQFRCRVGAY 60  
DB 17 LGAPPGSSSELSAFSAARTTPLEGTSEMAVTFDKYVNIIGDFDVATGQFRCRVGAY 76  
QY 61 FFSFTAGKAPHKSLSWLVNRNDEVQALAFDEQRPPGARRAASQSAMLQLDYGDPTWRL 120  
DB 77 FFSFTAGKAPHKSLSWLVNRNDEVQALAFDEQRPPGARRAASQSAMLQLDYGDPTWRL 136  
QY 121 HGAPHYALGAPGATFSGYLIVYAD 143  
DB 137 HGAPHYALGAPGATFSGYLIVYAD 159  
RESULT 8  
AAB61488  
ID AAB61488 standard; protein; 199 AA.  
AC AAB61488;  
XX  
XX 04-APR-2001 (first entry)  
DT  
DE Murine MANGO 245 protein.  
XX  
KW TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
KM autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
XX pancreatic; skeletal; muscle.  
OS Mus musculus.  
XX  
PN WO200100672-A1.  
XX  
PD 04-JAN-2001.  
XX  
PF 29-JUN-2000; 2000WO-US18184.  
XX  
PR 29-JUN-1999; 99US-0342687.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;  
XX  
DR WPI; 2001-050127/06.  
XX  
PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
PT disorders (e.g. jaundice) -  
XX  
PS Claim 1; Fig 29; 262pp; English.  
XX  
CC The present invention relates to cDNAs encoding TANGO 244,  
CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
CC The nucleic acids, proteins and protein modulators are useful for  
CC treating colonic disorders, inflammatory diseases, tumors,  
CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
CC allergic diseases, cardiovascular diseases, brain disorders,  
CC degenerative diseases placental, pancreatic, skeletal and muscle  
CC disorders.  
XX  
SQ Sequence 199 AA;  
Query Match 95.0%; Score 707.5; DB 22; Length 199;  
Best Local Similarity 95.8%; Pred. No. 1.7e-72;  
Matches 137; Conservative 3; Mismatches 2; Indels 1; Gaps 1;  
QY 1 LGPTPGSSSELSAFSAARTTPLEGTSEMAVTFDKYVNIIGDFDVATGQFRCRVGAY 60  
DB 17 LGAPA-GPGSSSELSAFSAARTTPLEGTSEMAVTFDKYVNIIGDFDVATGQFRCRVGAY 75  
QY 61 FFSFTAGKAPHKSLSWLVNRNDEVQALAFDEQRPPGARRAASQSAMLQLDYGDPTWRL 120  
DB 76 FFSFTAGKAPHKSLSWLVNRNDEVQALAFDEQRPPGARRAASQSAMLQLDYGDPTWRL 135





PD	04-JAN-2001.
XX	
PF	29-JUN--2000; 2000MO-US18184.
XX	
PR	29-JUN-1999; 99US-0342687.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
PI	Holtzman DA, Barnes TM, Frazer CC, Sharp JD;
XX	
DR	WPI; 2001-050127/06.
PT	Isolated cDNA encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and
PT	TANGO 245 proteins, useful in the treatment of inflammatory diseases
PT	(e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver
PT	disorders (e.g. jaundice) -
PS	
XX	Disclosure: Fig 28; 262pp; English.
CC	The present invention relates to cDNAs encoding TANGO 244,
CC	TANGO 246, TANGO 275, TANGO 300 and TANGO 245 proteins.
CC	The nucleic acids, proteins and protein modulators are useful for
CC	treating colonic disorders, inflammatory diseases, tumors,
CC	renal disorders, liver disorders, lung disorders, autoimmune diseases,
CC	allergic diseases, cardiovascular diseases, brain disorders,
CC	degenerative diseases placental, pancreatic, skeletal and muscle
CC	disorders.
XX	
SQ	Sequence 126 AA:
	Query Match 83.5%; Score 622; DB 22; Length 126;
	Best Local Similarity 95.2%; Pred. No. 5.6e-63;
	Matches 120; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY	15 AFSARTTPELGTESEMAVTFKRVYVNTGGDPDVAITGCFRCVPGAVFFSFTAGAPPKSL 74
DB	1 AFSARTTPLEGASEMAVTFPDKRVVNTGGDFDAATGOFRCRVPAGVFFSPVGRAPHKSL 60
QY	75 SVMIVRNNDVEQALAFPEORPPGARBARAASOSAMQLDTGYDTVMRLKGAPHVALGAPCAT 134
DB	61 SVMIVRNNDVEQALAFDQRPPSARRAASOSAMQLDTGYDTVMRLHGAPVALGAPCAT 120
QY	135 FSGYLTV 140
DB	121 FSGYLTV 126
RESULT 13	
AAB42189	
ID	AAB42189 standard; Protein; 123 AA.
AC	AAB42189;
XX	
DT	08-FEB-2001 (first entry)
DE	
XX	Human ORFX ORF1953 polypeptide sequence SEQ ID NO:3906.
XX	
KM	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KM	vulnerary; antipsoptic; antiparkinsonian; nootropic; neuroprotective;
KM	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KM	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KM	hypotensive; dermatological; immunosuppressive; antiinflammatory;
KM	antiviral; antibacterial; antifungal; antihematic; antithyroid;
KM	antihaemic; gene therapy; cancer; proliferative disorder; hypertension;
KM	neurodegenerative disorder; osteoarthritis; graft vs host disease;
KM	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KM	cholesterol ester storage; systemic lupus erythematosus; infection;
KM	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KM	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KM	bone damage; cartilage damage; antiinflammatory disease; coagulation;
KM	Thromboside; contraceptive.
XS	
XS	Homo sapiens.
XS	

XX PN WO200058473-A2.  
XX PD 05-OCT-2000.  
XX PF 31-MAR-2000; 2000MO-US08621.  
XX PR 31-MAR-1999; 99US-0127607.  
XX PR 02-APR-1999; 99US-0127636.  
XX PR 05-APR-1999; 99US-0127728.  
XX PR 30-MAR-2000; 2000US-0540763.  
XX (CURA-) CURAGEN CORP.  
XX PA Shimkete RA, Leach M;  
XX PI WPI: 2000-602362/57.  
XX DR N-PSDB; AAC76398.  
XX PT Novel nucleic acids and peptides derived from open reading frame X,  
XX PT useful for treating e.g. cancers, proliferative disorders,  
XX PT neurodegenerative disorders and cardiovascular disease -  
XX PS Claim 11; Page 3051-3052; 5507pp; English.  
XX CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
XX CC which represent the human ORF open reading frames 1 to 3161. The ORF  
XX CC sequences have activities such as: cytostatic; hepatocytic; vulnary;  
XX CC antiproliferative; antiparkinsonian; nootropic; neuroprotective;  
XX CC osteopathic; anticonvulsant; antitachycardic; immunosuppressive;  
XX CC immunostimulant; cardiant; thrombolytic; coagulant; vasodilator;  
XX CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
XX CC antitumour; antibacterial; antiviral; antifungal; antitumour;  
XX CC antitumour; and antitumour. The sequences can be used for determining  
XX CC the presence of or predisposition to, or preventing or treating  
XX CC pathological conditions associated with an ORF-associated disorder. The  
XX CC nucleic acids can be used to express ORF proteins in gene therapy  
XX CC vectors. The proteins and nucleic acids may be used to treat cancers,  
XX CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
XX CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
XX CC hyperension, hypothyroidism, cholesterol ester storage, systemic lupus  
XX CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
XX CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
XX CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
XX CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
XX CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX SQ Sequence 123 AA;  
XX Query Match 83.0%; Score 618; DB 21; Length 123;  
XX Best Local Similarity 98.4%; Pred. No. 1.6e-62;  
XX Matches 121; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
XX  
XX QY 8 GSSESRSAFSAARTPLEGTSEMAVTFPKVYVYVIGDGDVATGQRCRCPGAVFFSFAG 67  
XX DB 1 GSSESRSAFSAARTPLEGTSEMAVTFPKVYVYVIGDGDVATGQRCRCPGAVFFSFAG 60  
XX QY 68 KAPKSLSVMLVRNDEVQALAFDQRPGARRAASQASAMQLDYGDTVMRLHGAPHYA 127  
XX DB 61 KAPKSLSVMLVRNDEVQALAFDQRPGARRAASQASAMQLDYGDTVMRLHGAPHYA 120  
XX QY 128 LGA 130  
XX DB 121 LGA 123  
XX  
XX RESULT 14  
XX ABG70383 standard; Protein; 221 AA.  
XX AC ABG70383;  
XX DT 05-NOV-2002 (first entry)

XX DE Adipocyte complement-related C1q Tumour Necrosis Factor-like protein #1.  
XX KW Human; NOX; NOX-associated disorder; cardiomyopathy; atherosclerosis;  
XX KW cell signal processing; metabolic pathway modulation; metabolic disorder;  
XX KW obesity; diabetes; infectious disease; neurodegenerative disorder; acne;  
XX KW Alzheimer's disease; Parkinson's disease; immune disorder; cancer;  
XX KW haematopoietic disorder; cirrhosis; pancreatitis; learning defect;  
XX KW memory defect; infertility; congenital heart defect; hair growth;  
XX KW pigmentation disorder; endocrine disorder; respiratory disease; health;  
XX KW gastro-intestinal disease; reproductive; neurological disease;  
XX KW bone marrow transplantation; endocrine disease; allergy; inflammation;  
XX KW nephrological disorder; urinary system disorder; age-related disorder;  
XX KW neuropsychiatric disorder; EGF-related protein; SCUBA1; TEN-14;  
XX KW adipocyte complement-related C1q tumour necrosis factor; out at first;  
XX KW beta adrenergic receptor kinase; BphA/ehk-2; glucose transporter;  
XX KW type Ia membrane sushi-containing domain; butyrophilin;  
XX KW type Ia membrane-sushi domain containing.  
XX OS Homo sapiens.  
XX PN WO200257453-A2.  
XX PD 25-JUL-2002.  
XX PF 19-DEC-2001; 2001WO-US0331.  
XX PR 19-DEC-2000; 2000US-265704P.  
XX PR 20-DEC-2000; 2000US-257314P.  
XX PR 02-MAY-2001; 2001US-288153P.  
XX PR 29-MAY-2001; 2001US-294075P.  
XX PR 24-JUL-2001; 2001US-307506P.  
XX PR 10-AUG-2001; 2001US-311590P.  
XX PR 10-AUG-2001; 2001US-311613P.  
XX PR 29-AUG-2001; 2001US-315617P.  
XX PR 14-SEP-2001; 2001US-322358P.  
XX PA (CURA-) CURAGEN CORP.  
XX PI Gangoli EA, Paturajan M, Vernet CM, Malvankar UM, Kekuda R;  
XX PI Stone DJ, Anderson D, Shimkete RA, Burgess CE, Zernusen BD, Liu X;  
XX PI Sytek KA, Casman SJ, Boldog FL, Smithson G, Li L, Ji W;  
XX DR WPI: 2002-590744/63.  
XX DR N-PSDB; ABS52095.  
XX PT Novel isolated NOX polypeptide useful for treating cardiomyopathy,  
XX PT atherosclerosis, metabolic disorders, diabetes, obesity, infectious  
XX PT disease, anorexia, neurodegenerative disorders, Alzheimer's disease or  
XX PT cancer -  
XX PS Claim 1; Page 28; 318pp; English.  
XX  
XX CC The present invention relates to new NOX polypeptides. The invention is  
XX CC useful for treating or preventing a NOX-associated disorder such as  
XX CC cardiomyopathy or atherosclerosis, where the disorder is related to cell  
XX CC signal processing and metabolic pathway modulation in a subject.  
XX CC preferably human. The invention is also useful for treating metabolic  
XX CC disorders (e.g. obesity), diabetes, infectious disease, neurodegenerative  
XX CC disorders (e.g. Alzheimer's disease, Parkinson's disease), immune  
XX CC disorders, haematopoietic disorders and various cancers. The molecules of  
XX CC the invention are also useful for treating or preventing cirrhosis,  
XX CC pancreatitis, learning and memory defects, infertility, congenital heart  
XX CC defects, acne, hair growth, pigmentation disorders, endocrine disorders,  
XX CC respiratory disease, gastro-intestinal diseases, reproductive diseases,  
XX CC neurological diseases, bone marrow transplantation, endocrine diseases,  
XX CC allergy and inflammation, nephrological disorders, urinary system  
XX CC disorders, neuropsychiatric disorders and age-related disorders.  
XX CC The present amino acid sequence represents a NOX protein of the  
XX CC invention.  
XX SQ Sequence 221 AA;

Query Match 60.1%; Score 448; DB 23; Length 221;  
 Best Local Similarity 66.0%; Pred. No. 8.5e-43;  
 Matches 95; Conservative 10; Mismatches 31; Indels 8; Gaps 3;

QY 5 PGPGSSSLRSASFSAARTTPELGTSSEMAVTFPKVYVNIIGDQDVATGQFRCRVPGAYFFSF 64  
 DB 66 PGDSSAMRSAFSAARTTPELGTSSEMAVTFPKVYVNIIGDQDVATGQFRCRVPGAYFFSF 125

QY 65 TAGKAPPKSLSVMLVRRNDEVQALAFDE--QRRPGARRASQSMQLQDYGDTWLRRLH- 121  
 DB 126 TLGKLPKRLSVKLMKRRDEVQAMLYDQASRR--REMOSQSYMLALRRGDVAVMLLSH 182

QY 122 --GAPHYALGAPGATFSGLYVYAD 143  
 DB 183 HDGYGAYSNHGKXITFSGFLVYPD 206

## RESULT 15

AAB61468  
 ID AAB61468 standard; protein; 130 AA.

AC AAB61468;

DT 04-APR-2001 (first entry)

DE Mature monkey MANGO 245 protein.

XX TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
 KW autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
 KW pancreatic; skeletal; muscle.

OS Catarhini sp.

XX WO200100672-A1.

XX 04-JAN-2001.

PF 29-JUN-2000; 2000WO-US18184.

XX 29-JUN-1999; 99US-0342687.

XX (MILL-) MILLENNIUM PHARM INC.

PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD,

DR WPI; 2001-050127/06.

XX Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
 PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
 PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
 XX disorders (e.g. jaundice) -

PS Disclosure; Page 243; 262pp; English.

XX The present invention relates to cDNAs encoding TANGO 244,

CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.

CC The nucleic acids, proteins and protein modulators are useful for

CC treating colonic disorders, inflammatory diseases, tumors,

CC renal disorders, liver disorders, lung disorders, autoimmune diseases,

CC allergic diseases, cardiovascular diseases, brain disorders,

CC degenerative diseases placental, pancreatic, skeletal and muscle

CC disorders.

XX Sequence 130 AA;

Query Match 57.2%; Score 426.5; DB 22; Length 130;

Best Local Similarity 62.1%; Pred. No. 1.2e-40;

Matches 90; Conservative 2; Mismatches 20; Indels 33; Gaps 2;

QY 2 GPTPGSSSLRSASFSAARTTPELGTSSEMAVTFPKVYVNIIGDQDVATGQFRCRVPGAYF 61  
 DB 1 GPAPGSSSLRSASFSAARTTPELGASSEMAVTFPKVYVNIIGDQDVATGQFRCRVPGAYF 60

QY 62 FSFTAGKAPPKSLSVMLVRRNDEVQALAFDEQRRPGARRASQSMQLQDYGDTWLRRLH 121  
 DB 61 FSFTVGAPPKSLSVMLALRR-----GDAVWLISH 90

QY 122 ---GAPHYALGAPGATFSGLYVYAD 143  
 DB 91 DHGYGAYSNHGKXITFSGFLVYPD 115

Search completed: January 12, 2004, 08:15:35  
 Job time : 21.5499 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:15:44 ; Search time 17.412 Seconds

(Without alignments)  
1655.023 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_17\_159

Sequence: 1 LGPTPGGSSSELSARSAFSAAR.....PHYALGAPGATFSGYLIVAD 143

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747907 seqs, 201509753 residues

Total number of hits satisfying chosen parameters: 747907

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	745	100.0	329	12	US-10-085-167-2
2	739	99.2	329	12	US-10-236-055A-14
3	706.5	94.8	326	12	US-10-236-055A-16
4	222	29.8	202	15	US-10-091-458A-43
5	222	29.8	420	12	US-10-236-055A-2
6	222	29.8	746	10	US-09-738-973-185
7	222	29.8	746	10	US-09-854-133-185
8	222	29.8	746	15	US-10-144-649A-185
9	222	29.8	908	12	US-10-309-422-10
10	222	29.8	909	12	US-10-309-422-12
11	222	29.8	957	12	US-10-309-422-14
12	222	29.8	958	12	US-10-309-422-26
13	222	29.8	992	12	US-10-309-422-18
14	222	29.8	993	12	US-10-309-422-30
15	222	29.8	1042	12	US-10-309-422-8

16	222	29.8	1043	12	US-10-309-422-20	Sequence 20, Appl
17	222	29.8	1091	12	US-10-309-422-12	Sequence 12, Appl
18	222	29.8	1092	12	US-10-309-422-24	Sequence 24, Appl
19	222	29.8	1126	12	US-10-309-422-16	Sequence 16, Appl
20	222	29.8	1127	12	US-10-309-422-28	Sequence 28, Appl
21	216	29.0	409	12	US-10-236-055A-4	Sequence 4, Appl
22	200	26.8	294	11	US-09-866-050A-294	Sequence 294, App
23	200	26.8	294	12	US-10-236-055A-8	Sequence 8, Appl
24	197	26.4	285	11	US-09-866-050A-382	Sequence 382, App
25	196	26.3	285	16	US-10-234-000-5	Sequence 5, Appl
26	196	26.3	330	12	US-10-236-055A-6	Sequence 6, Appl
27	193	25.9	146	12	US-10-325-717-24	Sequence 24, Appl
28	193	25.9	153	12	US-10-325-717-58	Sequence 58, Appl
29	192	25.8	163	12	US-10-325-717-13	Sequence 23, Appl
30	192	25.8	163	12	US-10-325-717-54	Sequence 54, Appl
31	192	25.8	163	12	US-10-325-717-55	Sequence 55, Appl
32	192	25.8	163	12	US-10-325-717-59	Sequence 59, Appl
33	191	25.6	144	12	US-10-325-717-5	Sequence 5, Appl
34	191	25.6	144	12	US-10-325-717-25	Sequence 25, Appl
35	191	25.6	145	12	US-10-325-717-11	Sequence 11, Appl
36	191	25.6	145	12	US-10-325-717-28	Sequence 28, Appl
37	191	25.6	157	12	US-10-325-717-4	Sequence 4, Appl
38	191	25.6	163	12	US-10-325-717-10	Sequence 10, Appl
39	191	25.6	163	12	US-10-325-717-19	Sequence 19, Appl
40	191	25.6	187	12	US-10-325-717-13	Sequence 13, Appl
41	191	25.6	187	12	US-10-325-717-34	Sequence 34, Appl
42	191	25.6	193	12	US-10-325-717-12	Sequence 12, Appl
43	191	25.6	193	12	US-10-325-717-31	Sequence 31, Appl
44	191	25.6	203	12	US-10-325-717-3	Sequence 3, Appl
45	191	25.6	203	12	US-10-325-717-22	Sequence 22, Appl

#### ALIGNMENTS

RESULT 1  
US-10-085-167-2  
; Sequence 2, Application US/10085167  
; Publication No. US20030170781A1  
; GENERAL INFORMATION:  
; APPLICANT: Hollaway, James L.  
; APPLICANT: Lok, Si  
; TITLE OF INVENTION: SECRETED PROTEIN ZACRPA  
; FILE REFERENCE: 99-29  
; CURRENT APPLICATION NUMBER: US/10/085,167  
; CURRENT FILING DATE: 2002-02-27  
; PRIOR APPLICATION NUMBER: 60/141,928  
; PRIOR FILING DATE: 1999-07-01  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 329  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-085-167-2

Query Match	100.0%	Score 745;	DB 12;	Length 329;
Best Local Similarity	100.0%	Pred. No. 3.6e-78;		
Matches 143;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	LGPTPGGSSSELSARSAFSAARTTPELGTSMAVTFDKVYVNIIGDFDVATGQRCRVGAY	60	
DB	17	LGPTPGGSSSELSARSAFSAARTTPELGTSMAVTFDKVYVNIIGDFDVATGQRCRVGAY	76	
QY	61	FRSFTGKAPHGLSLVLRNDEVQYALAFDEQRGRGARRAASQSMQLDYGDTWML	120	
DB	77	FRSFTGKAPHGLSLVLRNDEVQYALAFDEQRGRGARRAASQSMQLDYGDTWML	136	
QY	121	HGAPHYALGAPGATFSGYLIVAD	143	
DB	137	HGAPHYALGAPGATFSGYLIVAD	159	

```
RESULT 2
US-10-236-055A-14
; Sequence 14, Application US/10236055A
; Publication No. US20030134328A1
; GENERAL INFORMATION:
; APPLICANT: Basham, Beth E.
; APPLICANT: Forsythe, Ian
; APPLICANT: Gorman, Daniel M.
; APPLICANT: Mattson, Jeanine
; APPLICANT: Moshrefi, Mehرداد
; APPLICANT: Patham, Christl
; TITLE OF INVENTION: MAMMALIAN GENES; RELATED REAGENTS
; FILE REFERENCE: DX01343K
; CURRENT APPLICATION NUMBER: US/10/236,055A
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/317,988
; PRIOR FILING DATE: 2001-09-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-236-055A-14

Query Match          99.2%; Score 739; DB 12; Length 329;
Best Local Similarity 99.3%; Pred. No. 1,8e-77;
Matches 142; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LGPTPGSSSELRSAFSAARTTPLEGTSEMAVTFDKYVYVIGDPDFVATGFCRCVPGAY 60
DB 17 LGPTPGSSSELRSAFSAARTTPLEGTSEMAVTFDKYVYVIGDPDFVATGFCRCVPGAY 76
QY 61 FFSTAGKAPHKSLSVMLVNRDEVOALAFDEQRPGARRAASQSMLOLDYGDYVWRL 120
DB 77 FFSTAGKAPHKSLSVMLVNRDEVOALAFDEQRPGARRAASQSMLOLDYGDYVWRL 136
QY 121 HGAPHYALGAPGATFSGYLVYAD 143
DB 137 HGAPHYALGAPGATFSGYLVYAD 159

RESULT 3
US-10-236-055A-16
; Sequence 16, Application US/10236055A
; Publication No. US20030134328A1
; GENERAL INFORMATION:
; APPLICANT: Basham, Beth E.
; APPLICANT: Forsythe, Ian
; APPLICANT: Gorman, Daniel M.
; APPLICANT: Mattson, Jeanine
; APPLICANT: Moshrefi, Mehرداد
; APPLICANT: Patham, Christl
; TITLE OF INVENTION: MAMMALIAN GENES; RELATED REAGENTS
; FILE REFERENCE: DX01343K
; CURRENT APPLICATION NUMBER: US/10/236,055A
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/317,988
; PRIOR FILING DATE: 2001-09-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 326
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-236-055A-16

Query Match          94.8%; Score 706.5; DB 12; Length 326;
Best Local Similarity 96.5%; Pred. No. 1.1e-73;
Matches 138; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 LGPTPGSSSELRSAFSAARTTPLEGTSEMAVTFDKYVYVIGDPDFVATGFCRCVPGAY 60
DB 17 LGPTPGSSSELRSAFSAARTTPLEGTSEMAVTFDKYVYVIGDPDFVATGFCRCVPGAY 76
```

```
DB 17 LGPA-GBGSSELRSAFSAARTTPLEGTSEMAVTFDKYVYVIGDPDFVATGFCRCVPGAY 75
QY 61 FFSTAGKAPHKSLSVMLVNRDEVOALAFDEQRPGARRAASQSMLOLDYGDYVWRL 120
DB 76 FFSTAGKAPHKSLSVMLVNRDEVOALAFDEQRPGARRAASQSMLOLDYGDYVWRL 135
QY 121 HGAPHYALGAPGATFSGYLVYAD 143
DB 136 HGAPHYALGAPGATFSGYLVYAD 159

RESULT 4
US-10-091-458-43
; Sequence 43, Application US/10091458
; Publication No. US20030068627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT206C1
; CURRENT APPLICATION NUMBER: US/10/091,458
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 09/764,900
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/225,757
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/226,868
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/216,647
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,267
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/216,880
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,270
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/251,869
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/235,834
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/234,274
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/234,223
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/228,924
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/224,518
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,369
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/224,519
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,964
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/241,809
; PRIOR FILING DATE: 2000-10-20
```

PRIOR APPLICATION NUMBER: 60/249,299  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/236,327  
 PRIOR FILING DATE: 2000-09-29  
 PRIOR APPLICATION NUMBER: 60/241,785  
 PRIOR FILING DATE: 2000-10-20  
 PRIOR APPLICATION NUMBER: 60/244,617  
 PRIOR FILING DATE: 2000-11-01  
 PRIOR APPLICATION NUMBER: 60/225,268  
 PRIOR FILING DATE: 2000-08-14  
 PRIOR APPLICATION NUMBER: 60/236,368  
 PRIOR FILING DATE: 2000-09-29  
 PRIOR APPLICATION NUMBER: 60/251,856  
 PRIOR FILING DATE: 2000-12-08  
 PRIOR APPLICATION NUMBER: 60/251,868  
 PRIOR FILING DATE: 2000-12-08  
 PRIOR APPLICATION NUMBER: 60/229,344  
 PRIOR FILING DATE: 2000-09-01  
 PRIOR APPLICATION NUMBER: 60/234,997  
 PRIOR FILING DATE: 2000-09-25  
 PRIOR APPLICATION NUMBER: 60/229,343  
 PRIOR FILING DATE: 2000-09-01  
 PRIOR APPLICATION NUMBER: 60/229,345  
 PRIOR FILING DATE: 2000-09-01  
 PRIOR APPLICATION NUMBER: 60/229,287  
 PRIOR FILING DATE: 2000-09-01  
 PRIOR APPLICATION NUMBER: 60/229,513  
 PRIOR FILING DATE: 2000-09-05  
 PRIOR APPLICATION NUMBER: 60/231,413  
 PRIOR FILING DATE: 2000-09-08  
 PRIOR APPLICATION NUMBER: 60/229,509  
 PRIOR FILING DATE: 2000-09-05  
 PRIOR APPLICATION NUMBER: 60/236,367  
 PRIOR FILING DATE: 2000-09-29  
 PRIOR APPLICATION NUMBER: 60/237,039  
 PRIOR FILING DATE: 2000-10-02  
 PRIOR APPLICATION NUMBER: 60/237,038  
 PRIOR FILING DATE: 2000-10-02  
 PRIOR APPLICATION NUMBER: 60/236,370  
 PRIOR FILING DATE: 2000-09-29  
 PRIOR APPLICATION NUMBER: 60/236,802  
 PRIOR FILING DATE: 2000-10-02  
 PRIOR APPLICATION NUMBER: 60/237,037  
 PRIOR FILING DATE: 2000-10-02  
 PRIOR APPLICATION NUMBER: 60/237,040  
 PRIOR FILING DATE: 2000-10-02  
 PRIOR APPLICATION NUMBER: 60/240,960  
 PRIOR FILING DATE: 2000-10-20  
 PRIOR APPLICATION NUMBER: 60/239,935  
 PRIOR FILING DATE: 2000-10-13  
 PRIOR APPLICATION NUMBER: 60/239,937  
 PRIOR FILING DATE: 2000-10-13  
 PRIOR APPLICATION NUMBER: 60/241,787  
 PRIOR FILING DATE: 2000-10-20  
 PRIOR APPLICATION NUMBER: 60/246,474  
 PRIOR FILING DATE: 2000-11-08  
 PRIOR APPLICATION NUMBER: 60/246,532  
 PRIOR FILING DATE: 2000-11-08  
 PRIOR APPLICATION NUMBER: 60/249,216  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,210  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/226,681  
 PRIOR FILING DATE: 2000-08-22  
 PRIOR APPLICATION NUMBER: 60/225,759  
 PRIOR FILING DATE: 2000-08-14  
 PRIOR APPLICATION NUMBER: 60/225,213  
 PRIOR FILING DATE: 2000-08-14  
 PRIOR APPLICATION NUMBER: 60/227,182  
 PRIOR FILING DATE: 2000-08-22  
 PRIOR APPLICATION NUMBER: 60/225,214  
 PRIOR FILING DATE: 2000-08-14  
 PRIOR APPLICATION NUMBER: 60/235,836

PRIOR FILING DATE: 2000-09-27  
 PRIOR APPLICATION NUMBER: 60/230,438  
 PRIOR FILING DATE: 2000-09-06  
 PRIOR APPLICATION NUMBER: 60/215,135  
 PRIOR FILING DATE: 2000-06-30  
 PRIOR APPLICATION NUMBER: 60/225,266  
 PRIOR FILING DATE: 2000-08-14  
 PRIOR APPLICATION NUMBER: 60/249,218  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,208  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,213  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,212  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,207  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,245  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,244  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,217  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,211  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,215  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,264  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,214  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,297  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/232,400  
 PRIOR FILING DATE: 2000-09-14  
 PRIOR APPLICATION NUMBER: 60/231,242  
 PRIOR FILING DATE: 2000-09-08  
 PRIOR APPLICATION NUMBER: 60/232,081  
 PRIOR FILING DATE: 2000-09-08  
 PRIOR APPLICATION NUMBER: 60/232,080  
 PRIOR FILING DATE: 2000-09-08  
 PRIOR APPLICATION NUMBER: 60/231,414  
 PRIOR FILING DATE: 2000-09-08  
 PRIOR APPLICATION NUMBER: 60/231,244  
 PRIOR FILING DATE: 2000-09-08  
 PRIOR APPLICATION NUMBER: 60/233,064  
 PRIOR FILING DATE: 2000-09-14  
 PRIOR APPLICATION NUMBER: 60/233,063  
 PRIOR FILING DATE: 2000-09-14  
 PRIOR APPLICATION NUMBER: 60/232,397  
 PRIOR FILING DATE: 2000-09-14  
 PRIOR APPLICATION NUMBER: 60/232,399  
 PRIOR FILING DATE: 2000-09-14  
 PRIOR APPLICATION NUMBER: 60/232,401  
 PRIOR FILING DATE: 2000-09-14  
 PRIOR APPLICATION NUMBER: 60/241,808  
 PRIOR FILING DATE: 2000-10-20  
 PRIOR APPLICATION NUMBER: 60/241,826  
 PRIOR FILING DATE: 2000-10-20  
 PRIOR APPLICATION NUMBER: 60/241,786  
 PRIOR FILING DATE: 2000-10-20  
 PRIOR APPLICATION NUMBER: 60/241,221  
 PRIOR FILING DATE: 2000-10-20  
 PRIOR APPLICATION NUMBER: 60/246,475  
 PRIOR FILING DATE: 2000-11-08  
 PRIOR APPLICATION NUMBER: 60/231,243  
 PRIOR FILING DATE: 2000-09-08

Query Match 29.8%; Score 222; DB 15; Length 202;  
 Best Local Similarity 41.7%; Pred. NO. 1.4e-17;  
 Matches 58; Conservative 18; Mismatches 59; Indels 4; Gaps 4;

7 PSSMRSAPSAATTPL-EGTSEMAVTFDKVYVNIIGDFVATGQFRCAVGAFFSFT 65

DB 66 PUPQMRVAFSAARTSNLAPGTLDQPIVFDLLNNIGETFDLQGRFNCVNGTYVFIH 125  
QY 66 AGK-APHKSLSYMLVNRNDEVQALAFDEQRRPGARRAASOSAMLQDYGDTWLRHGAP 124  
DB 126 MLKLAIVNVPLVYNLMKN-EVLVSAYANDGAPD-HETASNHAILQLFGQDQIWLRLHGA 183  
QY 125 HYALGAPGATFSGYLYAD 143  
DB 184 IYGSSWKYSTFSGYLYOD 202

## RESULT 5

US-10-236-055A-2  
Sequence 2, Application US/10236055A  
Publication No. US20030134328A1  
GENERAL INFORMATION:  
APPLICANT: Baaham, Beth E.  
APPLICANT: Forsythe, Ian  
APPLICANT: Gorman, Daniel M.  
APPLICANT: Mattoon, Jeanine  
APPLICANT: Moshrefi, Mehرداد  
APPLICANT: Parham, Christi  
TITLE OF INVENTION: MAMMALIAN GENES; RELATED REAGENTS  
FILE REFERENCE: DX01343K  
CURRENT APPLICATION NUMBER: US/10/236,055A  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/317,988  
PRIOR FILING DATE: 2001-09-06  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 2  
LENGTH: 420  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-236-055A-2

## Query Match

Best Local Similarity 41.7%; Score 222; DB 12; Length 420;  
Pred. No. 3.7e-17;  
Matches 58; Conservative 18; Mismatches 59; Indels 4; Gaps 4;

QY 7 PGSELRSAPFAATTP-L-EGTSEMAVTFDKYVNIIGDFVATGQRCRVPAGAYFFSFT 65  
DB 284 PUPQMRVAFSAARTSNLAPGTLDQPIVFDLLNNIGETFDLQGRFNCVNGTYVFIH 343  
QY 66 AGK-APHKSLSYMLVNRNDEVQALAFDEQRRPGARRAASOSAMLQDYGDTWLRHGAP 124  
DB 344 MLKLAIVNVPLVYNLMKN-EVLVSAYANDGAPD-HETASNHAILQLFGQDQIWLRLHGA 401  
QY 125 HYALGAPGATFSGYLYAD 143  
DB 402 IYGSSWKYSTFSGYLYOD 420

## RESULT 6

US-09-738-973-185  
Sequence 185, Application US/09738973  
Patent No. US20020110563A1  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Henderson, Robert A.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Flinn, Steven P.  
APPLICANT: Mohamach, Radodoh  
APPLICANT: Algate, Paul A.  
APPLICANT: Secrist, Heather  
APPLICANT: Indirias, Carol Yoseph  
APPLICANT: Benson, Darin R.  
APPLICANT: Elliott, Mark  
APPLICANT: Mannion, Jane  
APPLICANT: Kalos, Michael D.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER

FILE REFERENCE: 210121.475C9  
CURRENT APPLICATION NUMBER: US/09/738,973  
CURRENT FILING DATE: 2000-12-14  
NUMBER OF SEQ ID NOS: 587  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 185  
LENGTH: 746  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-738-973-185

## Query Match

Best Local Similarity 29.8%; Score 222; DB 10; Length 746;  
Pred. No. 8e-17;  
Matches 58; Conservative 18; Mismatches 59; Indels 4; Gaps 4;

QY 7 PGSELRSAPFAATTP-L-EGTSEMAVTFDKYVNIIGDFVATGQRCRVPAGAYFFSFT 65  
DB 610 PUPQMRVAFSAARTSNLAPGTLDQPIVFDLLNNIGETFDLQGRFNCVNGTYVFIH 669  
QY 66 AGK-APHKSLSYMLVNRNDEVQALAFDEQRRPGARRAASOSAMLQDYGDTWLRHGAP 124  
DB 670 MLKLAIVNVPLVYNLMKN-EVLVSAYANDGAPD-HETASNHAILQLFGQDQIWLRLHGA 727  
QY 125 HYALGAPGATFSGYLYAD 143  
DB 728 IYGSSWKYSTFSGYLYOD 746

## RESULT 7

US-09-854-133-185  
Sequence 185, Application US/09854133  
Publication No. US20020183499A1  
GENERAL INFORMATION:  
APPLICANT: Lodes, Michael J.  
APPLICANT: Mohamach, Radodoh  
APPLICANT: Henderson, Robert A.  
APPLICANT: Benson, Darin R.  
APPLICANT: Secrist, Heather  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
FILE REFERENCE: 210121.475C10  
CURRENT APPLICATION NUMBER: US/09/854,133  
CURRENT FILING DATE: 2001-05-11  
NUMBER OF SEQ ID NOS: 735  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 185  
LENGTH: 746  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-854-133-185

## Query Match

Best Local Similarity 29.8%; Score 222; DB 10; Length 746;  
Pred. No. 8e-17;  
Matches 58; Conservative 18; Mismatches 59; Indels 4; Gaps 4;

QY 7 PGSELRSAPFAATTP-L-EGTSEMAVTFDKYVNIIGDFVATGQRCRVPAGAYFFSFT 65  
DB 610 PUPQMRVAFSAARTSNLAPGTLDQPIVFDLLNNIGETFDLQGRFNCVNGTYVFIH 669  
QY 66 AGK-APHKSLSYMLVNRNDEVQALAFDEQRRPGARRAASOSAMLQDYGDTWLRHGAP 124  
DB 670 MLKLAIVNVPLVYNLMKN-EVLVSAYANDGAPD-HETASNHAILQLFGQDQIWLRLHGA 727  
QY 125 HYALGAPGATFSGYLYAD 143  
DB 728 IYGSSWKYSTFSGYLYOD 746

## RESULT 8

US-10-144-649A-185  
Sequence 185, Application US/10144649A  
Publication No. US20030118599A1  
GENERAL INFORMATION:



125 QY

QY 7 PGSSRLRSASFSAARTTPL-EGTSEMAVTFD

QY 7 PGSELSAFAAARTPL-EGTSEMAVTFDKVYVNI GGFDFVATGQFRCRVP GAYFFSFT 65

Db 821 PLPQGMRAFAAARKSNLA PGTLDQPIVFDDLNNLNGSTFPLQZGRFPCPNGYVFIFH 880  
 QY 66 AGK-A-PHKSLSVMTLYRNFDEVQALAFDEQRPGARRAASQAMQLDYGDTWIRLHGAP 124  
 Db 881 MLKLAIVNPPLYNNLMKN-EEVTVSAIYANDGAPD-HEIYSNHAIIQLFEGQDIIWRLHGA 938  
 QY 125 HYALGAPGATFSGYLVYAD 143  
 Db 939 IYGSWMKYSTFSGYLLYQD 957

```

RESULT 12
US-10-309-422-26
; Sequence 26, Application US/10309422
; Publication NO. US20030139587A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030139587A1 Human Proteins and Polynucleotides Encoding
; FILE REFERENCE: LEX-0142-USA
; CURRENT APPLICATION NUMBER: US/10/309,422
; CURRENT FILING DATE: 2002-12-03
; PRIOR APPLICATION NUMBER: US/09/798,771
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 60/186,557
; PRIOR FILING DATE: 2000-03-02
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 958
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-309-422-26

```

[illegible]

```

1 RESULT 13
2 US-10-309-422-18
3 ; Sequence 18, Application US/10/309422
4 ; Publication No. US20030139587A1
5 ; GENERAL INFORMATION:
6 ; APPLICANT: Waikie, D. Wade
7 ; APPLICANT: Wilganowski, Nathaniel L.
8 ; APPLICANT: Turner, C. Alexander Jr.
9 ; TITLE OF INVENTION: No. US20030139587A1 Human Proteins and Polynucleotides Encoding
10 ; FILE REFERENCE: LEX-0142-USA
11 ; CURRENT APPLICATION NUMBER: US/10/309,422
12 ; CURRENT FILING DATE: 2002-12-03
13 ; PRIOR APPLICATION NUMBER: US/09/798,771
14 ; PRIOR FILING DATE: 2001-03-02
15 ; PRIOR APPLICATION NUMBER: US 60/186,557
16 ; PRIOR FILING DATE: 2000-03-02
17 ; NUMBER OF SEQ ID NOS: 43
18 ; SOFTWARE: FastSeq for Windows Version 4.0
19 ; SEQ ID NO: 18
20 ; LENGTH: 992

```

	:	TYPE:	PRT	
	:	ORGANISM:	homo sapiens	
	:	US-10-309-422-18		
Query Match		29.8%	Score 222;	DB 12; Length 992;
Best Local Similarity		41.7%;	Pred. No. 1.2e-16;	
Matches 58;	Conservative 18;	Mismatches 59;	Indels 4;	Gaps 4.
QY	7	PGSSELRAFPAAATTPPL-EGTSEAAVVFDDKYVYNIGGDFDVAATGOEFCRVPGAAYFSESFT	65	
Dd	856	PLPQQMRARAFSAATISNLAPEGTLDQPIYFDLLLNLTGSTPTLQGRFNCPPNGTYVTFLEH	915	
QY	66	AGK-APEHKSLSVMLVRNDEVOALAFDEQRPFGARPAASQSAMLOLDYGDTWLRHLHGAP	124	
Dd	916	MLKLIAVNPLLVNNLMKN-EEVLSAAYANDGAPD-HETSMHAIIQLFOGDQIWLRLHRGA	973	
QY	125	HYALGAPGAEFGSYLVVAD	143	
Dd	974	IYGSWMKYSTSTSGYLITOD	992	

```

RESULT 14
US-10-309-422-30
; Sequence 30, Application US/10309422
; Publication No. US20030139587A1
GENERAL INFORMATION:
APPLICANT: Walke, D. Wade
APPLICANT: Wilganowski, Nathaniel L.
APPLICANT: Turner, C. Alexander Jr.
TITLE OF INVENTION: No. US20030139587A1 Human Proteins and Polynucleotides Encodind
FILE REFERENCE: LEX-0142-USA
CURRENT APPLICATION NUMBER: US/10/309,422
CURRENT FILING DATE: 2002-12-03
PRIOR APPLICATION NUMBER: US/09/798,771
PRIOR FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: US 60/186,557
PRIOR FILING DATE: 2000-03-02
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 993
; TYPE: prt
; ORGANISM: homo sapiens
US-10-309-422-30

```

	Query Match	29.8%	Score 222	DB 12	Length 993
	Best Local Similarity	41.7%	Pred No. 1,2e-16		
	Matches	58	Conservative	18	Mismatches 59; Indels 4; Gaps 4
OY	7	PGSSLRSAFSAARTPL - EGTSENAVTFDKYYNIGDFDVATGQFRCRVGAYFFSFT	65		
	:	:::::			
Dd	857	PLPQGMRVAFAARSRLNLAGTLDGPVFEDLLNLNGEFPDLQGRFCNPVGVTVFIFH	916		
OY	66	AGK -APHKSLSYMLVRNDEVOTALFDDQRPRGARRAASQSAMLDDVDGYTWLRLHGAP	124		
	:	:::::			
Dd	917	MUKLVANVPYLNYLMKN -EEVLVSAYAMDGAPD -HETASNAILQLFOGDQIMTLRIHGCA	974		
OY	125	HVALGAPGRTSGYLYVAD	143		
	:	:::::			
Dd	975	IYGSSWKYSTSEGILLYID	993		

```

: RESULT 15
: US-10-309-422-8
: Sequence 8, Application US/10309422
: Publication No. US20030139587A1
: GENERAL INFORMATION:
: APPLICANT: Walke, D. Wade
: APPLICANT: Wiganowski, Nathaniel L.
: APPLICANT: Turner, C. Alexander Jr.
: TITLE OF INVENTION: No. US20030139587A1 Human Proteins and Polynucleotides Encodinf
: FILE REFERENCE: LEX-0142-USA
: CURRENT APPLICATION NUMBER: US/10/309,422

```

;/ CURRENT FILING DATE: 2002-12-03  
;/ PRIOR APPLICATION NUMBER: US/09/798,771  
;/ PRIOR FILING DATE: 2001-03-02  
;/ PRIOR APPLICATION NUMBER: US 60/186,557  
;/ PRIOR FILING DATE: 2000-03-02  
;/ NUMBER OF SEQ ID NOS: 43  
;/ SOFTWARE: FastSeq for Windows Version 4.0  
;/ SEQ ID NO 8  
;/ LENGTH: 1042  
;/ TYPE: PRT  
;/ ORGANISM: homo sapiens  
US-10-309-422-8

## Query Match

29.8%; Score 222; DB 12; Length 1042;

Best Local Similarity: 41.7%; Pred. No. 1.2e-16;

Matches 58; Conservative 18; Mismatches 59; Indels 4; Gaps 4;

QY 7 FGSELRGAFSAARTTPL-EGTSEMAVTFEDKVVYVNIIGDFVATGQPRCRYGAYFFSFT 65  
| : : | | | | | : : : : : | : : | | | | |  
Db 906 PLPQOMRVAFSAARTSNLAPGTLDPIVFDLLNNLGETFDLQGRNCPVNGTYVFLFH 965  
QY 66 AGK-APKSLSVMLVRNDEVOALAFDEORRPGARRAASOSAMLQLDYDGTWTLRLHGAP 124  
| : : | | | | | : : : : : | : : | | | | |  
Db 966 MLKLAIVNVPVLYVNLKKN-EEVLVSAYANDGAPD-HETASNHAILQLFGDQIWLRLHRGA 1023  
QY 125 HYALGAFGATFSGYLVYAD 143  
| : : | | | | | : : : : : | : : | | | | |  
Db 1024 IYGSSWKYTFSGYLLYOD 1042

Search completed: January 12, 2004, 08:33:37  
Job time : 18.4112 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:14 ; Search time 7.27844 Seconds

(without alignments)  
831.284 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_17\_159

Perfect score: 745

Sequence: 1 LGRTPGSGSELRSAPSAAR.....PHYALGAPATFGSYLYAD 143

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PTCUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	222	29.8	746	4 US-09-370-838-185	Sequence 185, App
2	200	26.8	294	3 US-09-188-930-294	Sequence 294, App
3	200	26.8	294	4 US-09-312-283C-294	Sequence 294, App
4	197	26.4	285	4 US-09-312-283C-382	Sequence 382, App
5	191	25.6	231	4 US-09-530-423-2	Sequence 2, Appli
6	191	25.6	244	4 US-08-463-911-7	Sequence 7, Appli
7	191	25.6	244	3 US-09-140-804-3	Sequence 3, Appli
8	191	25.6	244	4 US-09-336-536-20	Sequence 20, Appli
9	191	25.6	244	4 US-09-530-423-1	Sequence 1, Appli
10	191	25.6	244	4 US-09-686-838B-3	Sequence 3, Appli
11	191	25.6	244	4 US-09-911-176B-48	Sequence 48, Appli
12	191	25.6	244	4 US-09-552-225A-3	Sequence 3, Appli
13	191	25.6	244	4 US-09-619-740-51	Sequence 51, Appli
14	191	25.6	244	4 US-09-776-976-6	Sequence 6, Appli
15	191	25.6	244	4 US-09-909-547-6	Sequence 6, Appli
16	191	25.6	244	4 US-09-569-852B-6	Sequence 6, Appli
17	189	25.4	228	4 US-09-336-536-4	Sequence 4, Appli
18	189	25.4	243	4 US-09-140-804-2	Sequence 2, Appli
19	189	25.4	243	4 US-09-336-536-3	Sequence 3, Appli
20	189	25.4	243	4 US-09-686-838B-2	Sequence 2, Appli
21	188	25.2	259	4 US-09-996-243-47	Sequence 47, Appli
22	188	25.2	260	4 US-09-489-847-198	Sequence 188, App
23	188	25.2	287	4 US-09-489-847-349	Sequence 349, App
24	185	24.8	247	4 US-08-463-911-2	Sequence 8, Appli
25	185	24.8	247	3 US-09-140-804-8	Sequence 3, Appli
26	185	24.8	247	3 US-09-118-408-3	Sequence 3, Appli
27	185	24.8	247	4 US-09-506-855-3	Sequence 3, Appli

28	185	24.8	247	4 US-09-686-838B-8	Sequence 8, Appli
29	185	24.8	247	4 US-09-911-176B-3	Sequence 3, Appli
30	185	24.8	247	4 US-09-619-740-3	Sequence 3, Appli
31	185	24.8	247	4 US-09-776-976-4	Sequence 4, Appli
32	185	24.8	247	4 US-09-506-852-3	Sequence 3, Appli
33	185	24.8	247	4 US-09-909-547-4	Sequence 4, Appli
34	179.5	24.1	243	3 US-09-188-930-295	Sequence 295, App
35	179.5	24.1	243	4 US-09-312-283C-295	Sequence 295, App
36	178.5	24.0	228	4 US-09-336-536-11	Sequence 11, Appli
37	178.5	24.0	228	4 US-09-336-536-10	Sequence 10, Appli
38	177	23.8	247	4 US-09-776-976-2	Sequence 2, Appli
39	177	23.8	247	4 US-09-909-547-2	Sequence 2, Appli
40	170.5	22.9	281	3 US-09-118-408-2	Sequence 2, Appli
41	170.5	22.9	281	4 US-09-506-855-2	Sequence 2, Appli
42	170.5	22.9	281	4 US-09-911-176B-2	Sequence 2, Appli
43	170.5	22.9	281	4 US-09-619-740-2	Sequence 2, Appli
44	170.5	22.9	281	4 US-09-506-852-2	Sequence 2, Appli
45	167.5	22.5	130	4 US-09-485-316A-11	Sequence 11, Appli

## ALIGNMENTS

```
RESULT 1
US-09-370-838-185
; Sequence 185, Application US/09370838
; Patent No. 644425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Monamach, Roadoh
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370, 838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 185
; LENGTH: 746
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-370-838-185

Query Match      29.8%; Score 222; DB 4; Length 746;
Best Local Similarity 41.7%; Pred. No. 2e-18;
Matches 58; Conservative 18; Mismatches 59; Indels 4; Gaps 4;

OY 7 PGSELRSAPSAARTPL-EGTSEMAVTPDKYVYNIGDPDVANGOPRCRVPGAYPSPFT 65
    ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 610 PLQOQKRVASAKRTSNLAGTLDOPVFDLLNNIGETDLDGRENCEVNGTYFIFH 669
    ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 66 AGK-APHKSTSVMLVNNRDEVQALAPDEQRPPGARRAASQAMLOLDYGTWMLRLGAP 124
    ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 670 MLALVAVVPLVYVLMKN-EELVLSAYANDAPD-HEFASVHAIITQIFGQQLVRLHRCGA 727
    ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 125 HYALGAPGATFGSYLYAD 143
    ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 728 IYGSWXYSTFGSYLYYOD 746
    ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 2
US-09-188-930-294
; Sequence 294, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
```

APPLICANT: Murison, James Greg  
TITLE OF INVENTION: Compositions Isolated From Skin Cells  
FILE REFERENCE: 11000.1011c1  
CURRENT APPLICATION NUMBER: US/09/186,930A  
CURRENT FILING DATE: 1998-11-09  
NUMBER OF SEQ ID NOS: 348  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 294  
LENGTH: 294  
TYPE: PRT  
ORGANISM: Rat  
US-09-186-930-294

Query Match 26.8%; Score 200; DB 3; Length 294;  
Best Local Similarity 34.9%; Pred. No. 2.7e-16;  
Matches 53; Conservative 30; Mismatches 45; Indels 24; Gaps 8;

QY 5 PGP---GSSSELSAFSA--RTTPEGTSEMAVTFDKYVNIIGDFVATGQFRCRVGA 59  
DB 147 PGPCCGSSRAKSAFSAVATKSYPR--RLPIKFDKILMNEGCHYNASSGKFCVSPGI 203  
QY 60 YFFSFTAGKAPHKSLSVMLVNRDEVOALAFDEQRPPGARRAASQSMLOLDYDGTWLR 119  
DB 204 YFTYDITLA-NKHLAIGLVHN-GQYRIRTFD--ANTGNHDVASGSTITLALKEGDEWLO 259  
QY 120 LHGA-----PHYALGAPGATFSGYLVYAD 143  
DB 260 IFYSEONGLFYDPYWT-----DSLFTGFLIYAD 287

RESULT 3  
US-09-312-283C-294  
Sequence 294; Application US/09312283C  
Patent No. 6573095  
GENERAL INFORMATION:  
APPLICANT: Watson, James D.  
APPLICANT: Strachan, Lorna  
APPLICANT: Sleeman, Matthew  
APPLICANT: Onrust, Rene  
APPLICANT: Murison, James G.  
APPLICANT: Kumble, Krishanand D.  
TITLE OF INVENTION: Compositions Isolated from Skin Cells  
FILE REFERENCE: 11000.1011c2  
CURRENT APPLICATION NUMBER: US/09/312,283C  
CURRENT FILING DATE: 1999-05-14  
NUMBER OF SEQ ID NOS: 425  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 294  
LENGTH: 294  
TYPE: PRT  
ORGANISM: Mouse  
US-09-312-283C-294

Query Match 26.8%; Score 200; DB 4; Length 294;  
Best Local Similarity 34.9%; Pred. No. 2.7e-16;  
Matches 53; Conservative 30; Mismatches 45; Indels 24; Gaps 8;

QY 5 PGP---GSSSELSAFSA--RTTPEGTSEMAVTFDKYVNIIGDFVATGQFRCRVGA 59  
DB 147 PGPCCGSSRAKSAFSAVATKSYPR--RLPIKFDKILMNEGCHYNASSGKFCVSPGI 203  
QY 60 YFFSFTAGKAPHKSLSVMLVNRDEVOALAFDEQRPPGARRAASQSMLOLDYDGTWLR 119  
DB 204 YFTYDITLA-NKHLAIGLVHN-GQYRIRTFD--ANTGNHDVASGSTITLALKEGDEWLO 259  
QY 120 LHGA-----PHYALGAPGATFSGYLVYAD 143  
DB 260 IFYSEONGLFYDPYWT-----DSLFTGFLIYAD 287

RESULT 4

US-09-312-283C-382  
Sequence 382; Application US/09312283C  
Patent No. 6573095  
GENERAL INFORMATION:  
APPLICANT: Watson, James D.  
APPLICANT: Strachan, Lorna  
APPLICANT: Sleeman, Matthew  
APPLICANT: Onrust, Rene  
APPLICANT: Murison, James G.  
APPLICANT: Kumble, Krishanand D.  
TITLE OF INVENTION: Compositions Isolated from Skin Cells  
FILE REFERENCE: 11000.1011c2  
CURRENT APPLICATION NUMBER: US/09/312,283C  
CURRENT FILING DATE: 1999-05-14  
NUMBER OF SEQ ID NOS: 425  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 382  
LENGTH: 285  
TYPE: PRT  
ORGANISM: Mouse  
US-09-312-283C-382

Query Match 26.4%; Score 197; DB 4; Length 285;  
Best Local Similarity 34.2%; Pred. No. 6e-16;  
Matches 52; Conservative 31; Mismatches 45; Indels 24; Gaps 8;

QY 5 PGP---GSSSELSAFSA--RTTPEGTSEMAVTFDKYVNIIGDFVATGQFRCRVGA 59  
DB 138 PGPCCGSSRAKSAFSAVATKSYPR--RLPIKFDKILMNEGCHYNASSGKFCVSPGI 194  
QY 60 YFFSFTAGKAPHKSLSVMLVNRDEVOALAFDEQRPPGARRAASQSMLOLDYDGTWLR 119  
DB 195 YFTYDITLA-NKHLAIGLVHN-GQYRIRTFD--ANTGNHDVASGSTITLALKEGDEWLO 250  
QY 120 LHGA-----PHYALGAPGATFSGYLVYAD 143  
DB 251 IFYSEONGLFYDPYWT-----DSLFTGFLIYAD 278

RESULT 5  
US-09-530-423-2  
Sequence 2; Application US/09530423  
Patent No. 6461821  
GENERAL INFORMATION:  
APPLICANT: Otsuka Pharmaceutical Co., Ltd.  
TITLE OF INVENTION: Smooth muscle growth inhibitory composition, a  
TITLE OF INVENTION: diagnostic method for arteriosclerosis and a kit  
FILE REFERENCE: P98-51  
CURRENT APPLICATION NUMBER: US/09/530,423  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: JP H9-297569  
PRIOR FILING DATE: 1997-10-29  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 231  
TYPE: PRT  
ORGANISM: Abdominal fat tissue from myoma uteri  
US-09-530-423-2

Query Match 25.6%; Score 191; DB 4; Length 231;  
Best Local Similarity 31.8%; Pred. No. 2.4e-15;  
Matches 47; Conservative 29; Mismatches 54; Indels 18; Gaps 6;

QY 5 PGPSSSELSAFSAARTTPEGTSEMAVTFDKYVNIIGDFVATGQFRCRVGAFFSF 64  
DB 91 PGEAGYVYRSASFVGLFETVYI-INMPIRFKIRYNOONHYDGSIGKFKHCNIPGLIYFAY 149  
QY 65 TAGKAPH-----KSLSVMLVNRDEVOALAFDEQRPPGARRAASQSMLOLDYDGTWLR 119  
DB 150 -----HITVYMKOVKSLFK-KDKAMLFTYDYOENNVDQ-ASGSVILHLEVGQWLO 201

QY 120 LHGAPH----YALGAPGATFSGYLTVAD 143  
Db 202 VYGEGERNGLYADNDNDSTFTGFLLYHD 229

## RESULT 6

US-08-463-911-7  
Sequence 7, Application US/08463911  
Patent No. 5869330  
GENERAL INFORMATION:  
APPLICANT: Scherer, Philipp E.  
APPLICANT: Lodish, Harvey F.  
TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED  
TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCTES  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
City: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/463,911  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: WH195-05  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 244 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-463-911-7

Query Match 25.6%; Score 191; DB 2; Length 244;  
Best Local Similarity 31.8%; Pred. No. 2.6e-15;

Matches 47; Conservative 29; Mismatches 54; Indels 18; Gaps 6;

QY 5 PGGGSELRSAPSAARTTPELGTSMAVTFDKYVNIIGDFVATGQFCRVGAYFSGF 64  
Db 104 PEGGAVVYRSAPSVGLETVT-IPNMPIRFTKIFVQONHYDSTGKFCNIPGLYFAY 162  
QY 65 TAGKAPH-----KSLVMLVRNDEVQALAFDEORRPGARRAASOSAMQLDYGDTVMR 119  
Db 163 -----HITVYMKDVVSLFK-KDKAMLFYDQYQENNVDQ-ASGSVLLHLEVGQVWLQ 214  
QY 120 LHGAPH----YALGAPGATFSGYLTVAD 143  
Db 215 VYGEGERNGLYADNDNDSTFTGFLLYHD 242

## RESULT 7

US-09-140-804-3  
Sequence 3, Application US/09140804  
Patent No. 6197930  
GENERAL INFORMATION:  
APPLICANT: Shepard, Paul O.  
APPLICANT: Humes, Jacqueline M.  
TITLE OF INVENTION: ADIPOCTE-SPECIFIC PROTEIN HOMOLOGS  
FILE REFERENCE: 97-49

CURRENT APPLICATION NUMBER: US/09/140,804  
CURRENT FILING DATE: 1998-08-26  
EARLIER APPLICATION NUMBER: 60/056,983  
EARLIER FILING DATE: 1997-08-26  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 3  
LENGTH: 244  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-140-804-3

Query Match 25.6%; Score 191; DB 3; Length 244;  
Best Local Similarity 31.8%; Pred. No. 2.6e-15;

Matches 47; Conservative 29; Mismatches 54; Indels 18; Gaps 6;

QY 5 PGGGSELRSAPSAARTTPELGTSMAVTFDKYVNIIGDFVATGQFCRVGAYFSGF 64  
Db 104 PEGGAVVYRSAPSVGLETVT-IPNMPIRFTKIFVQONHYDSTGKFCNIPGLYFAY 162  
QY 65 TAGKAPH-----KSLVMLVRNDEVQALAFDEORRPGARRAASOSAMQLDYGDTVMR 119  
Db 163 -----HITVYMKDVVSLFK-KDKAMLFYDQYQENNVDQ-ASGSVLLHLEVGQVWLQ 214  
QY 120 LHGAPH----YALGAPGATFSGYLTVAD 143  
Db 215 VYGEGERNGLYADNDNDSTFTGFLLYHD 242

## RESULT 8

US-09-336-536-20  
Sequence 20, Application US/09336536  
Patent No. 6406884  
GENERAL INFORMATION:  
APPLICANT: Leiby, K.  
APPLICANT: McKay, C.  
APPLICANT: Bossone, S.  
TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF  
FILE REFERENCE: 7853-144  
CURRENT APPLICATION NUMBER: US/09/336,536  
CURRENT FILING DATE: 1999-06-18  
NUMBER OF SEQ ID NOS: 75  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 20  
LENGTH: 244  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-336-536-20

Query Match 25.6%; Score 191; DB 4; Length 244;  
Best Local Similarity 31.8%; Pred. No. 2.6e-15;

Matches 47; Conservative 29; Mismatches 54; Indels 18; Gaps 6;

QY 5 PGGGSELRSAPSAARTTPELGTSMAVTFDKYVNIIGDFVATGQFCRVGAYFSGF 64  
Db 104 PEGGAVVYRSAPSVGLETVT-IPNMPIRFTKIFVQONHYDSTGKFCNIPGLYFAY 162  
QY 65 TAGKAPH-----KSLVMLVRNDEVQALAFDEORRPGARRAASOSAMQLDYGDTVMR 119  
Db 163 -----HITVYMKDVVSLFK-KDKAMLFYDQYQENNVDQ-ASGSVLLHLEVGQVWLQ 214  
QY 120 LHGAPH----YALGAPGATFSGYLTVAD 143  
Db 215 VYGEGERNGLYADNDNDSTFTGFLLYHD 242

## RESULT 9

US-09-530-423-1  
Sequence 1, Application US/09530423  
Patent No. 6461821  
GENERAL INFORMATION:  
APPLICANT: Otsuka Pharmaceutical Co., Ltd.  
TITLE OF INVENTION: Smooth muscle growth inhibitory composition, a

```

; TITLE OF INVENTION: diagnostic method for arteriosclerosis and a kit
; FILE REFERENCE: 98-51
; CURRENT APPLICATION NUMBER: US/09/530,423
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: JP H9-297569
; PRIOR FILING DATE: 1997-10-29
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Abdominal fat tissue from myoma uteri
US-09-530-423-1

```

```

Query Match          25.6%; Score 191; DB 4; Length 244;
Best Local Similarity 31.8%; Pred. No. 2.6e-15;
Matches 47; Conservative 29; Mismatches 54; Indels 18; Gaps 6;

```

```

QY 5 PGGSSSELRSASFSAARTTPLEGTSEMAVTFDKYYVNIIGDPVATGQRCRVGAYFFSF 64
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 104 PGEAGAYVYRSASFVGLETYVT-IPNMPIRFTKIFVNOQNHVDSGTGKHCNIPGLYFYAY 162
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 65 TAGKAPH-----KSLVWLVRNDEVQALAFDEQRPGARRAASQASAMLOLDYGDYVWL 119
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 163 -----HITVYMKDVKVSLEK-KDKAMLFITYDQYQENNVDQ-ASGSVLLHLEVGQVWLQ 214
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 120 LHGAPH-----YALGAPGATFSGYLYVAD 143
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 215 VYGEGERNGLYADNDNDSTFTGFLYHD 242
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 10
US-09-686-838B-3
; Sequence 3, Application US/09686838B
; Patent No. 6482612
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: Adipocyte-Specific Protein Homologs
; FILE REFERENCE: 97-4901
; CURRENT APPLICATION NUMBER: US/09/686,838B
; CURRENT FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/140,804
; PRIOR FILING DATE: 1998-08-26
; PRIOR APPLICATION NUMBER: US 60/056,983
; PRIOR FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-686-838B-3

```

```

Query Match          25.6%; Score 191; DB 4; Length 244;
Best Local Similarity 31.8%; Pred. No. 2.6e-15;
Matches 47; Conservative 29; Mismatches 54; Indels 18; Gaps 6;

```

```

QY 5 PGGSSSELRSASFSAARTTPLEGTSEMAVTFDKYYVNIIGDPVATGQRCRVGAYFFSF 64
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 104 PGEAGAYVYRSASFVGLETYVT-IPNMPIRFTKIFVNOQNHVDSGTGKHCNIPGLYFYAY 162
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 65 TAGKAPH-----KSLVWLVRNDEVQALAFDEQRPGARRAASQASAMLOLDYGDYVWL 119
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 163 -----HITVYMKDVKVSLEK-KDKAMLFITYDQYQENNVDQ-ASGSVLLHLEVGQVWLQ 214
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 120 LHGAPH-----YALGAPGATFSGYLYVAD 143
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 215 VYGEGERNGLYADNDNDSTFTGFLYHD 242
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 11

```

US-09-911-176B-48
; Sequence 48, Application US/09911176B
; Patent No. 6518403
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: ANTIBODIES THAT BIND AN
; FILE REFERENCE: 97-3001
; CURRENT APPLICATION NUMBER: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOG
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: US/09/911,176B
; PRIOR FILING DATE: 09/118,408
; PRIOR FILING DATE: 1998-07-17
; PRIOR APPLICATION NUMBER: 60/053,154
; PRIOR FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 48
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-911-176B-48

```

```

Query Match          25.6%; Score 191; DB 4; Length 244;
Best Local Similarity 31.8%; Pred. No. 2.6e-15;
Matches 47; Conservative 29; Mismatches 54; Indels 18; Gaps 6;

```

```

QY 5 PGGSSSELRSASFSAARTTPLEGTSEMAVTFDKYYVNIIGDPVATGQRCRVGAYFFSF 64
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 104 PGEAGAYVYRSASFVGLETYVT-IPNMPIRFTKIFVNOQNHVDSGTGKHCNIPGLYFYAY 162
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 65 TAGKAPH-----KSLVWLVRNDEVQALAFDEQRPGARRAASQASAMLOLDYGDYVWL 119
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 163 -----HITVYMKDVKVSLEK-KDKAMLFITYDQYQENNVDQ-ASGSVLLHLEVGQVWLQ 214
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 120 LHGAPH-----YALGAPGATFSGYLYVAD 143
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 215 VYGEGERNGLYADNDNDSTFTGFLYHD 242
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 12
US-09-552-225A-3
; Sequence 3, Application US/09552225A
; Patent No. 6521233
; GENERAL INFORMATION:
; APPLICANT: Piddington, Christopher S.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOG ZACRP3
; FILE REFERENCE: 99-09
; CURRENT APPLICATION NUMBER: US/09/552,225A
; CURRENT FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 60/130,199
; PRIOR FILING DATE: 1999-04-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-552-225A-3

```

```

Query Match          25.6%; Score 191; DB 4; Length 244;
Best Local Similarity 31.8%; Pred. No. 2.6e-15;
Matches 47; Conservative 29; Mismatches 54; Indels 18; Gaps 6;

```

```

QY 5 PGGSSSELRSASFSAARTTPLEGTSEMAVTFDKYYVNIIGDPVATGQRCRVGAYFFSF 64
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 104 PGEAGAYVYRSASFVGLETYVT-IPNMPIRFTKIFVNOQNHVDSGTGKHCNIPGLYFYAY 162
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 65 TAGKAPH-----KSLVWLVRNDEVQALAFDEQRPGARRAASQASAMLOLDYGDYVWL 119
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 163 -----HITVYMKDVKVSLEK-KDKAMLFITYDQYQENNVDQ-ASGSVLLHLEVGQVWLQ 214
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 120 LHGAPH-----YALGAPGATFSGYLYVAD 143
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



Db 215 VYGEGERNGLYADNDNDSTFTGFLYHD 242

RESULT 13  
US-09-619-740-51  
Sequence 51, Application US/09619740  
Patent No. 6544946  
GENERAL INFORMATION:  
APPLICANT: Sheppard, Paul O.  
APPLICANT: Laaser, Gerald W.  
APPLICANT: Bishop, Paul D.  
TITLE OF INVENTION: INHIBITORS FOR USE IN HEMOSTASIS AND IMMUNE FUNCTION  
FILE REFERENCE: 99-12C3  
CURRENT APPLICATION NUMBER: US/09/619,740  
CURRENT FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 09/253,604  
PRIOR FILING DATE: 1999-02-19  
PRIOR APPLICATION NUMBER: 09/444,794  
PRIOR FILING DATE: 1999-11-22  
PRIOR APPLICATION NUMBER: 09/506,855  
PRIOR FILING DATE: 2000-02-17  
NUMBER OF SEQ ID NOS: 55  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 51  
LENGTH: 244  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-619-740-51

Query Match 25.6%; Score 191; DB 4; Length 244;  
Best Local Similarity 31.8%; Pred. No. 2.6e-15;  
Matches 47; Conservative 29; Mismatches 54; Indels 18; Gaps 6;

Qy 5 PGGSSELRSASFSAARTTPEGTSEMAVTFDKYVNIIGDFVATGQRCRVPQAYFFSF 64  
Db 104 PGEGAVYRSASFVGLTETVT-IPNMPIRFTKIFVQNHNDSTGKFNHIGLYFFAY 162

Qy 65 TAGKAPH-----KSLSVMLVRNDEVOALAFDEQRPGARRASQASAMLDYGDYTWLR 119  
Db 163 -----HITVYMKDVKVSLEFK-KDKAMLFYDQYQENNVDO-ASGSVLLHLEVGDOVWLQ 214

Qy 120 LHGAPH-----YALGAPGATPSGYLYAD 143  
Db 215 VYGEGERNGLYADNDNDSTFTGFLYHD 242

RESULT 14  
US-09-776-976-6  
Sequence 6, Application US/09776976  
Patent No. 6566332  
GENERAL INFORMATION:  
APPLICANT: Fruehls, Joachim  
APPLICANT: Erickson, Mary Ruth  
APPLICANT: Yen, Frances  
APPLICANT: Blahin, Bernard  
TITLE OF INVENTION: OBG3 Globular Head and Uses Thereof for Decreasing Body Mass  
FILE REFERENCE: 76.US4.REG  
CURRENT APPLICATION NUMBER: US/09/776,976  
CURRENT FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 09/758,055  
PRIOR FILING DATE: 2001-01-10  
PRIOR APPLICATION NUMBER: US 60/176,228  
PRIOR FILING DATE: 2000-01-14  
PRIOR APPLICATION NUMBER: US 60/198,087  
PRIOR FILING DATE: 2000-04-13  
PRIOR APPLICATION NUMBER: US 60/299,881  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: Patent.pm  
SEQ ID NO 6  
LENGTH: 244  
TYPE: PRT

ORGANISM: Homo sapiens  
US-09-776-976-6

Query Match 25.6%; Score 191; DB 4; Length 244;  
Best Local Similarity 31.8%; Pred. No. 2.6e-15;  
Matches 47; Conservative 29; Mismatches 54; Indels 18; Gaps 6;

Qy 5 PGGSSELRSASFSAARTTPEGTSEMAVTFDKYVNIIGDFVATGQRCRVPQAYFFSF 64  
Db 104 PGEGAVYRSASFVGLTETVT-IPNMPIRFTKIFVQNHNDSTGKFNHIGLYFFAY 162

Qy 65 TAGKAPH-----KSLSVMLVRNDEVOALAFDEQRPGARRASQASAMLDYGDYTWLR 119  
Db 163 -----HITVYMKDVKVSLEFK-KDKAMLFYDQYQENNVDO-ASGSVLLHLEVGDOVWLQ 214

Qy 120 LHGAPH-----YALGAPGATPSGYLYAD 143  
Db 215 VYGEGERNGLYADNDNDSTFTGFLYHD 242

RESULT 15  
US-09-909-547-6  
Sequence 6, Application US/09909547  
Patent No. 6579852  
GENERAL INFORMATION:  
APPLICANT: Fruehls, Joachim  
APPLICANT: Erickson, Mary Ruth  
APPLICANT: Yen, Frances  
APPLICANT: Blahin, Bernard  
TITLE OF INVENTION: OBG3 Globular Head and Uses Thereof for Decreasing Body Mass  
FILE REFERENCE: 76.US6.CIP  
CURRENT APPLICATION NUMBER: US/09/909,547  
CURRENT FILING DATE: 2001-07-19  
PRIOR APPLICATION NUMBER: US 09/776,976  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 09/758,055  
PRIOR FILING DATE: 2001-01-10  
PRIOR APPLICATION NUMBER: US 60/299,881  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: US 60/198,087  
PRIOR FILING DATE: 2000-04-13  
PRIOR APPLICATION NUMBER: US 60/176,228  
PRIOR FILING DATE: 2000-01-14  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: Patent.pm  
SEQ ID NO 6  
LENGTH: 244  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-909-547-6

Query Match 25.6%; Score 191; DB 4; Length 244;  
Best Local Similarity 31.8%; Pred. No. 2.6e-15;  
Matches 47; Conservative 29; Mismatches 54; Indels 18; Gaps 6;

Qy 5 PGGSSELRSASFSAARTTPEGTSEMAVTFDKYVNIIGDFVATGQRCRVPQAYFFSF 64  
Db 104 PGEGAVYRSASFVGLTETVT-IPNMPIRFTKIFVQNHNDSTGKFNHIGLYFFAY 162

Qy 65 TAGKAPH-----KSLSVMLVRNDEVOALAFDEQRPGARRASQASAMLDYGDYTWLR 119  
Db 163 -----HITVYMKDVKVSLEFK-KDKAMLFYDQYQENNVDO-ASGSVLLHLEVGDOVWLQ 214

Qy 120 LHGAPH-----YALGAPGATPSGYLYAD 143  
Db 215 VYGEGERNGLYADNDNDSTFTGFLYHD 242

Search completed: January 12, 2004, 08:20:04  
Job time : 7.27844 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:09 ; Search time 18.1178 Seconds  
(without alignments)  
1661.397 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_17\_329

Perfect score: 1639

Sequence: 1 LGPTPGSGSELRSASFASAR.....LVYPDLAPAPPGIGASSEL 313

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR 76:\*

1: P1r1:\*

2: P1r2:\*

3: P1r3:\*

4: P1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	213	13.0	219	2	T14782
2	192	11.7	244	2	JC4708
3	184	11.2	215	2	B48150
4	182	11.1	253	1	CIH0B
5	175	10.7	744	1	S23298
6	173	10.6	744	1	A34246
7	172	10.5	245	1	CIH0C
8	170	10.4	744	2	S15435
9	169.5	10.3	674	2	S23297
10	168.5	10.3	253	2	S49158
11	168.5	10.3	680	2	S31216
12	166.5	10.2	680	1	CGH0D
13	165.5	10.1	743	1	S23779
14	163.5	10.0	674	2	S13301
15	163	9.9	224	2	A60032
16	160.5	9.8	246	2	S29328
17	159	9.7	245	2	S19018
18	158.5	9.7	253	2	I49560
19	156.5	9.5	193	2	A37873
20	150	9.2	196	2	A48150
21	150	9.2	245	1	CIH0A
22	144.5	8.8	215	2	C48150
23	144	8.8	1228	2	A57384
24	139.5	8.5	635	2	A57131
25	139	8.5	423	2	A55797
26	135.5	8.3	170	2	B57131
27	131.5	8.0	992	2	T08772
28	96.5	5.9	476	2	T29083
29	94	5.7	810	2	B84185

30	94	5.7	913	2	AB2587	translation initia
31	94	5.7	913	2	B97369	translation initia
32	94	5.7	1045	2	A39199	endoglucanase B (E
33	94	5.7	3938	2	T42761	Bassoon protein -
34	92.5	5.6	420	2	E75130	hypothetical prote
35	91.5	5.6	13055	2	T16580	hypothetical prote
36	91	5.6	882	2	S41034	hypothetical prote
37	90.5	5.5	2124	2	H83357	probable non-ribos
38	90	5.5	268	2	D75574	hypothetical prote
39	90	5.5	366	2	T52655	thioisulfate sulfur
40	90	5.5	606	2	S13367	Om(1D) protein - I
41	90	5.5	1239	1	Q0B810	BOLFI protein - hu
42	90	5.5	1433	2	A46053	bullous pemphigoid
43	90	5.5	7463	2	T36248	CPA peptide synthe
44	89.5	5.5	400	2	E75299	tellurium resistan
45	89	5.4	239	2	T36995	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

T14782  
hypothetical protein DKFZp586B0621.1 - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 02-Sep-2000

C:Accession: T14782  
R:Ottensmeyer, B.; Obermaier, B.; Mewes, H. W.; Gaassenhuber, U.; Wiemann, S.

submitted to the Protein Sequence Database, August 1999

A:Reference number: Z18184

A:Accession: T14782

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-219 <OTT>

A:Cross-references: EMBL:AL110261

A:Experimental source: adult uterus; clone DKFZp586B0621

C:Genetics:

A>Note: DKFZp586B0621.1

C:Superfamily: complement C1q carboxyl-terminal homology

Query Match 13.0%; Score 213; DB 2; Length 219;

Best Local Similarity 34.9%; Pred. No. 2.3e-10;

Matches 66; Conservative 18; Mismatches 83; Indels 22; Gaps 6;

QY	122	GAPHYALGAPGATFGC---YLVYADAD---ADAPAPGAPAP-----ERRSAFSARTR	169
DB	33	GAP-----GAPGEGGEGGPGPLPGPRGDPGPRGEGGPGAGGECVPPRSASFSAKRS	88
QY	170	SLVGSDDGPPRRHQPLAFDTEFVNIGSDPDAAGVFERCRLPGAYFFSFTLGKLPKXTLSV	229
DB	89	SRV----PPSPDAPLPFDRLVLYNEQGHYDAVTGKFTQVPGVYTPA-VHATVYRASLQF	142
QY	230	KLMKRDVEQAMTYDDGASRRREMOSQVMALARGDAWLLSHDHGAGYNSHGXYIT	289
DB	143	DIVVNGESIASFFQFGGWPKRPASLSGAVVRLEPEDQVWVGVDVYIGYVASIKXTDST	202
QY	290	FSGFVLYPD 298	
DB	203	FSGFVLYSD 211	

##### RESULT 2

JC4708  
gelatin-binding 28K protein precursor - human

N:Alternate names: adipose specific collagen-like factor

C:Species: Homo sapiens (man)

C>Date: 10-May-1996 #sequence\_revision 19-Jul-1996 #text\_change 20-Sep-1999

C:Accession: JC4708, J04944

R:Maeda, K.; Okubo, K.; Shimomura, T.; Funahashi, T.; Matsuzawa, Y.; Matsubara, K.

Biochem. Biophys. Res. Commun. 221, 286-289, 1996

A:Title: cDNA cloning and expression of a novel adipose specific collagen-like factor, af

A:Reference number: JC4708; MUID:96224171; PMID:8619847

A:Accession: JC4708

A: Molecule type: mRNA  
A: Residues: 1-244 <MAE>  
A: Cross-references: DDBJ:D45371; NID:G871886; PIDN:BA08227.1; PID:G871887  
A: Experimental source: adipose tissue  
R: Nakano, Y.; Tohe, T.; Choi-Miyura, N.H.; Mazda, T.; Tomita, M.  
J: Biochem. J. 120, 803-812, 1996  
A: Title: Isolation and characterization of G8P28, a novel gelatin-binding protein purified from rat adipose tissue  
A: Reference number: JC4944; MUID:97103474; PMID:8947845  
A: Accession: JC4944  
A: Molecule type: protein  
A: Residues: 19-38; 93-100; 101-112; 135-149; 173-178 <NAK>  
C: Comment: This protein is an endogenous factor that binds with a collagen-like domain.  
C: Genetics:  
A: Gene: apm1  
C: Superfamily: unassigned collagens; complement C1q carboxyl-terminal homology  
C: Keywords: adipose tissue; glycoprotein; hydroxyproline  
F: 1-18/Domain: signal sequence #status predicted <SIG>  
F: 19-244/Product: gelatin-binding 28kDa protein #status experimental <MAT>  
F: 142-107/Region: collagen-like  
F: 114-241/Domain: complement C1q carboxyl-terminal homology <C1Q>  
F: 95/Modified site: 4-hydroxyproline (Pro) #status experimental  
F: 230/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 11.7%; Score 192; DB 2; Length 244;  
Best Local Similarity 31.3%; Pred. No. 1.4e-08;  
Matches 47; Conservative 30; Mismatches 55; Indels 18; Gaps 6;

QY 5 PGPSSSLRSAPSAARTPLEGTSSEMAVTEDKVVYVNIIGDPDVAITGQFRCPGAPYFSPF 64  
Db 104 PGEAGVYVRSAPSVGLLETVYT-IPNMPRIFFKIFVQCNHNGDSGTGKPHCNIPGLYYEAY 162  
QY 65 TAGCAAPH-----KSLSVLVNRPDEVQALADEQRPPARAAASQASMLQIDYDGTWLR 119  
Db 163 -----HITVYMKVQVKSLEFR-KDKAMLEFDYQVQENNVQD-ASGSVLHLLEVDQVWLQ 214  
QY 120 LHGAPH-----YALGAPGATFGYLVTYADAD 145  
Db 215 VYGEGERNGLYADNDNDSTFTGFLIYHPTN 244

RESULT 3  
B48150  
hibernation-related protein HP-25 precursor - Siberian chipmunk  
C: Species: Eutamias sibiricus (Siberian chipmunk)  
C: Date: 16-Feb-1994 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
C: Accession: B48150; B41752  
R: Takamatsu, N.; Ohba, K.; Kondo, J.; Kondo, N.; Shiba, T.  
J: Takamatsu, N.; Ohba, K.; Kondo, J.; Kondo, N.; Shiba, T.  
Mol. Cell. Biol. 13, 1516-1521, 1993  
A: Title: Hibernation-associated gene regulation of plasma proteins with a collagen-like  
A: Reference number: A48150; MUID:93180798; PMID:8441393  
A: Accession: B48150  
A: Molecule type: mRNA; protein  
A: Residues: 1-215 <TKA>  
A: Cross-references: GB:012975; NID:G287469; PIDN:BA02352.1; PID:G287470  
A: Note: the source is designated as Tamias asiaticus in Genbank entry TMSHP25, release 1  
A: Note: sequence extracted from NCBI Backbone (NCBIN:125946, NCBI:P:125947)  
J: R. Biol. Chem. 267, 473-478, 1992  
J: Kondo, N.; Kondo, J.  
A: Title: Identification of novel blood proteins specific for mammalian hibernation.  
A: Reference number: A41752; MUID:92112696; PMID:1730610  
A: Accession: B41752  
A: Status: Preliminary  
A: Molecule type: protein  
A: Residues: 29-62; 84-130; 172-183; 187-192; 201-215 <KON>  
C: Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal homology  
C: Keywords: glycoprotein; hibernation; plasma  
F: 1-28/Domain: signal sequence #status predicted <SIG>  
F: 29-215/Product: hibernation-related protein HP-25 #status experimental <MAT>  
F: 91-215/Domain: complement C1q carboxyl-terminal homology <C1Q>  
F: 167/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 11.2%; Score 184; DB 2; Length 215;  
Best Local Similarity 30.4%; Pred. No. 5.6e-08;

F:33-116/Domain: collagenous, triple helix <COL>  
 F:328-249/Domain: complement C1q carboxyl-terminal homology <C1Q>  
 F:328/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimental  
 F:31/Disulfide bonds: interchain (to chain A-26) #status experimental  
 F:35,38,41,53,56,65,86,101,104,107/Modified site: 4-hydroxyproline (Pro) #status experimental  
 F:59,62,77,92,98,110/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F:59,62,98,110/Binding site: carboxylate (Lys) (covalent) #status experimental

Query Match 11.1%; Score 182; DB 1; Length 253;  
 Best Local Similarity 29.1%; Pred. No. 1e-07;  
 Matches 48; Conservative 32; Mismatches 61; Indels 24; Gaps 5;

149 PARGPAPRPP-----RSASARTSLVSDAGPGRHPLAFDFEFVIG 195  
 |||||  
 97 PKGGPAPAGAGPKGESGDYATOKIAPSAITTNV-----PLRRDQTIKFDHVTIMN 150  
 |||||  
 196 GDPDAAGVFCRLPGAYFFSFTLGKLPKRLSVLTKMKRDEVQAMT-YDDGASRRRMSQ 254  
 ::::|  
 151 NNYEPKSGKFTCKYKPGLYFTYHSS--KGNLCVNLWGRERAKVVFCDYAVNTFGVT 208  
 ::::|  
 255 SQSVMLARRGDAVWLSHDHGCGAVSNHGKTYTFSGFLVYPLD 299  
 ::::|  
 209 TCGVVLKLECGENVFLOQTDRN--SLMGEGANSIFSGFLFPDM 251  
 ::::|

RESULT 5  
 523298  
 collagen alpha 1(VIII) chain - chicken  
 C:Species: Gallus gallus (chicken)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: S23298  
 R:Ninomiya, Y.; Castagnola, P.; Gerecke, D.; Gordon, M.K.; Jacenko, O.; LuValle, P.; McMaguchi, N.; Olsen, B.R.  
 in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp.79-114, Academic Press  
 A:Title: The molecular biology of collagens with short triple-helical domains.  
 A:Reference number: S22243  
 A:Accession: S23298  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-744 <NN>  
 C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology <C1Q>  
 F:617-743/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 10.7%; Score 175; DB 1; Length 744;  
 Best Local Similarity 25.3%; Pred. No. 1.4e-06;  
 Matches 56; Conservative 28; Mismatches 77; Indels 60; Gaps 8;

120 LHGAPHALGAPGATFSGYLVYADADADAPA-RGPPAPRPPRS----- 161  
 |||||  
 540 LHGPP-----GKPGA-----LGPGQGPGLPGRPPGPPGPPAPVPTPAPQGEYLPDMG 589  
 |||||  
 162 -----AFSARTSLVSDAGPQ-----PRHOPLAFDFEFVIG 196  
 ::::|  
 590 LGIDGVTPHAYAKK-----GKNGGPAYEMPAPFAELTAPFPVGAIRKFDRLLYNQRQ 644  
 ::::|  
 197 DFDAAAGVFCRLPGAYFFSFTLGKLPKRLSVLTKMKRDEVQAMT-YDDGASRRRMSQ 256  
 ::::|  
 645 NYNPQTGIFTCVGVYFAFVH-HCKGNVWVALLFKNNNEPM-YTYDEVYKKGFLDQASG 702  
 ::::|  
 257 SVMALRRGDAVWLSHDHGCGAVSNHGKTYTFSGFLVYPLD 297  
 ::::|  
 703 SAVLLRPGDRVFLQMPSEQAAGLYAGQYHSSFGYLLYP 743  
 ::::|

RESULT 6  
 A34246  
 collagen alpha 1(VIII) chain precursor - rabbit  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: A34246  
 R:Yamauchi, N.; Benya, P.D.; van der Rest, M.; Ninomiya, Y.  
 J. Biol. Chem. 264, 16022-16029, 1989  
 A:Title: The cloning and sequencing of alpha1(VIII) collagen cDNAs demonstrate that type

omains similar to those of type X collagen.  
 A:Reference number: A34246; MUID:69380199; PMID:2476437  
 A:Accession: A34246  
 A:Molecule type: mRNA  
 A:Residues: 1-744 <YMM>  
 A:Cross-references: GB:05042; NID:g164895; PIDN:AAA31204.1; PID:g164896  
 C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology <C1Q>  
 F:1-30/Domain: signal sequence #status predicted <SIG>  
 F:21-744/Product: collagen alpha 1(VIII) chain #status predicted <MAT>  
 F:21-117/Region: amino-terminal nonhelical  
 F:118-571/Region: interrupted helical  
 F:572-744/Region: carboxyl-terminal nonhelical  
 F:617-743/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 10.6%; Score 173; DB 1; Length 744;  
 Best Local Similarity 25.3%; Pred. No. 2.1e-06;  
 Matches 56; Conservative 27; Mismatches 78; Indels 60; Gaps 8;

120 LHGAPHALGAPGATFSGYLVYADADADAPA-RGPPAPRPPRS----- 161  
 |||||  
 540 LHGPP-----GKPGA-----LGPGQGPGLPGRPPGPPGPPAPVPTPAPQGEYLPDMG 589  
 |||||  
 162 -----AFSARTSLVSDAGPQ-----PRHOPLAFDFEFVIG 196  
 ::::|  
 590 LGIDGVTPHAYAKK-----GKNGGPAYEMPAPFAELTAPFPVGAIRKFDRLLYNQRQ 644  
 ::::|  
 197 DFDAAAGVFCRLPGAYFFSFTLGKLPKRLSVLTKMKRDEVQAMT-YDDGASRRRMSQ 256  
 ::::|  
 645 NYNPQTGIFTCVGVYFAFVH-HCKGNVWVALLFKNNNEPM-YTYDEVYKKGFLDQASG 702  
 ::::|  
 257 SVMALRRGDAVWLSHDHGCGAVSNHGKTYTFSGFLVYPLD 297  
 ::::|  
 703 SAVLLRPGDRVFLQMPSEQAAGLYAGQYHSSFGYLLYP 743  
 ::::|

RESULT 7  
 C1HUC  
 complement subcomponent C1q chain C precursor - human  
 N:Alternate names: complement subcomponent C1q gamma chain  
 C:Species: Homo sapiens (man)  
 C:Date: 22-May-1991 #sequence\_revision 31-May-1996 #text\_change 22-May-1998  
 C:Accession: S14351; A03207  
 R:Sellar, G.C.; Blake, D.J.; Reid, K.B.M.  
 Biochem. J. 274, 481-490, 1991  
 A:Title: Characterization and organization of the genes encoding the A-, B- and C-chains  
 A:Reference number: S14350; MUID:91174759; PMID:1706597  
 A:Accession: S14351  
 A:Status: not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-245 <SEL>  
 R:Reid, K.B.M.  
 Biochem. J. 179, 367-371, 1979  
 A:Title: Complete amino acid sequences of the three collagen-like regions present in sub  
 A:Reference number: A90304; MUID:80020137; PMID:486087  
 A:Accession: A03207  
 A:Molecule type: protein  
 A:Residues: 29-56, 'P', 58-65, 'K', 67-71, 'P', 73-83, 'K', 85-86, 'D', 88-89, 'N', 91-122 <REI>  
 C:Comment: The first component of complement is a calcium-dependent complex of three  
 iation of C1r (enzyme), C1s (proenzyme), and the other eight components of complement.  
 C:Comment: The C1q subcomponent is composed of nine subunits, six of which are disulfide-  
 dimers of the C chain. Equimolar amounts of the A, B, and C chains are found after reduc  
 C:Genetics:  
 A:Gene: GDB:C1QG  
 A:Cross-references: GDB:128132; OMIM:120575  
 A:Map position: 1p36.3-1p34.1  
 A:Introns: 60/3  
 C:Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal hom  
 C:Keywords: complement pathway; glycoprotein; homodimer; hydroxylysine; hydroxyproline; I  
 F:1-28/Domain: signal sequence #status predicted <SIG>  
 F:29-245/Product: complement subcomponent C1q chain B #status predicted <MAT>  
 F:31-114/Domain: collagenous, triple helix <COL>  
 F:121-244/Domain: complement C1q carboxyl-terminal homology <C1Q>  
 F:332/Disulfide bonds: interchain #status experimental

F:36.39.42.45.54.63.81.93.96.99.105/Modified site: 4-hydroxyproline (Pro) #status experimental  
F:57.72.75/Modified site: 5-hydroxylysine (Lys) #status experimental  
F:75/Binding site: carbohydrate (Lys) (covalent) #status experimental

Query Match 10.5%; Score 172; DB 1; Length 245;

Best Local Similarity 23.8%; Pred. No. 6.4e-07; Mismatches 84; Indels 98; Gaps 13;

Matches 67; Conservative 33; Mismatches 84; Indels 98; Gaps 13;

QY 70 PHKSLVNLVNRNDEVQALFDEQRPPGARRASQSAQLQDLYGDTWLRLGHAHYALG 129  
DB 9 PHLGKLLTL-----LTLALRLGQANTGC-----YG-----IRMGGLP-G 43  
QY 130 ARGATFSG--LVYADADADAPA-----RGPPAP----- 157  
DB 44 APGK--DGTDLGPKGEGIPALPGIRPKQKQKBPGLPGPKGKGMPPGMPGVP 101  
QY 158 -----EP-----RSAPSAARTSLVSGDAGPGRHQP-----LAPTEFVNIG 196  
DB 102 MGIPPEPGEGRYKQKFGSFTVTR-----QTHQPPAPNSLIRFNAVLTPQG 149  
QY 197 PDFAAGVFRCLPGAYFPSTFLGKLPKRTLSVKLMKRNDEVQAMTYDDGASRRRMO 256  
DB 150 DDTSTGKFTCKVPGLYFVYHAS---HTANLCVLLYRSYGVKVTFCGHTSKTQVNSG 205  
QY 257 SWMLALRGDAVWLISHDHGCGAYSNHGKITYTSGFLVYP 238  
DB 206 GVLRLQVGEVWLAVNDY--YDMVGIGQSDSVFSGFLLPD 245

RESULT 8

S15435 collagen alpha 1(VIII) chain precursor - human

C/Species: Homo sapiens (man)

C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 13-Aug-1999

C/Accession: S15435

R/Mutagen: Y.; Matcel, M.G.; Yamaguchi, N.; Olsen, B.R.; Nimmiya, Y.

R/J Biochem 197, 615-622, 1991

A/Title: The complete primary structure of the human alpha-1(VIII) chain and assignment

A/Reference number: S15435; MUID:91231001; PMID:2029694

A/Accession: S15435

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-744 <MUR>

A/Cross-references: EMBL:X57527; NID:g30081; PIDN:CAA40748.1; PID:g30082

C/Genetics:

A/Gene: GDB:COL8A1

A/Cross-references: GDB:128104; OMIM:120251

C/Map position: 3q11.1-3q13.2

C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-744/Product: collagen alpha 1(VIII) chain #status predicted <MAT>

F:118-571/Region: amino-terminal nonhelical

F:572-744/Region: carboxyl-terminal nonhelical

F:617-743/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 10.4%; Score 170; DB 2; Length 744;

Best Local Similarity 25.9%; Pred. No. 3.7e-06; Mismatches 56; Conservative 27; Mismatches 83; Indels 50; Gaps 9;

QY 120 LHGAHYALGARGATSSGLVYADADADAPA--RGPPAPPEPR----- 161  
DB 540 LHGPP-----GKPGA-----LGPQGPGLPGPPGPPGPPAPVMPPTPPGGEYLPMG 589  
QY 162 -AFSAARTSLVSGDAGP--GPRHOPLAEDTE-----FVNIG-----DFDA 201  
DB 590 LGIDGVKPRHATGAKKKGKGRGAYEMPAFTALTAEPFVGGRVKNKLLVNGRQVNPQ 649  
QY 202 AGVPRCLRGAYFPSTFLGKLPKRTLSVKLMKRNDEVQAMTYDDGASRRRMO 261  
DB 650 TGIFCEVGVGVYFAVHV--HCKGQVWVVALPKNNEPVM--YTYDEVKKGFLDASGSAVL 707  
QY 262 LRRGDAVWLISHDHGCGAYSNHGKITYTSGFLVYP 237

DB 708 LRPGRVFLQMPRQAGLVAGQVHSSFGSYLLYP 743

RESULT 9

S23297 collagen alpha 1(X) chain precursor - chicken

N/Alternate names: type X collagen

C/Species: Gallus gallus (chicken)

C/Date: 07-Oct-1994 #sequence\_revision 10-Nov-1995 #text\_change 13-Aug-1999

C/Accession: S23297; A31896; S65594; S77711; I50218

R/Nimmiya, Y.; Caetano, P.; Gerecke, D.; Gordon, M.K.; Jacenko, O.; Luvalle, P.; McGe

guchi, N.; Olsen, B.R.

A/Title: The molecular biology of collagens with short triple-helical domains.

A/Reference number: S22243

A/Accession: S23297

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-674 <NIN>

R/Luvalle, P.; Nimmiya, Y.; Rosenblum, N.D.; Olsen, B.R.

J. Biol. Chem. 263, 18378-18385, 1988

A/Title: The type X collagen gene. Intron sequences split the 5'-untranslated region and

A/Reference number: A31896; MUID:89054019; PMID:2461368

A/Accession: A31896

A/Molecule type: mRNA

A/Residues: 1-75 <LUV>

R/Nimmiya, Y.; Gordon, M.; van der Rest, M.; Schmid, T.; Linemayer, T.; Olsen, B.R.

J. Biol. Chem. 261, 5041-5050, 1986

A/Title: The developmentally regulated type X collagen gene contains a long open reading

A/Reference number: I50218; MUID:86168227; PMID:3082876

A/Accession: S65594

A/Molecule type: DNA

A/Residues: 'T', '9', 'D', '11-12', 'EDOMKTYLFTM', '30-31', 'TKSGRAFTYWLQNVADLVSSHT', '48-89', 'L', '5

69', 'PQAVLSISWRTIKGSSCOIQNMVSIPLNMTLLSQVSYLLKSNIPITMS' <NIN1>

A/Cross-references: EMBL:M13496; NID:g211699; PIDN:AAA48736.1; PID:g211700

A/Accession: S77711

A/Molecule type: protein

A/Residues: 104-112, 'X', '114-117', '453-466' <NIN2>

C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

C/Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer; hydroxyproline;

F:1-18/Domain: signal sequence #status predicted <SIG>

F:547-673/Domain: complement C1q carboxyl-terminal homology <C1Q>

F:453-456/Modified site: hydroxyproline (Pro) #status experimental

F:611/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 10.3%; Score 169.5; DB 2; Length 674;

Best Local Similarity 25.3%; Pred. No. 3.6e-06; Mismatches 58; Conservative 21; Mismatches 75; Indels 75; Gaps 7;

QY 129 GARGATSSGLVYADADADAPA-----RGPPAPPEPR----- 160  
DB 454 GARG-----AKGAGAPGLPGPAGIATKGLGPMGPQPPGPKNSGEPGLPPG 504  
QY 161 -----SAPSAARTSLVSG-----SDAGPGRHOP 185  
DB 505 PPGPGGSTRIBGYKSGRSLSGSPKAKAGANQILTMPASATVILSKAYPGR-VPI 563  
QY 186 AFDTEFVNIGDPPDAAGVFRCLRGAYFPSTFLGKLPKRTLSVKLMKRNDEVQAMTYDD 245  
DB 564 KFDKILVNRQGHYDRTGIFTCRIPIGLYFYSYHV--HAKGTWVWVALYKNGSPVM--YTYDE 621  
QY 246 GASRRRMOQSWMLALRGDAVWLISHDHGCGAYSNHGKITYTSGFL 294  
DB 622 YQKGYLDOASGSAVIDLMENDQVWLQPLNPSNSNGLYSSSEYVHSSFGFL 670

RESULT 10

S49158 complement protein C1q beta chain precursor - rat

C/Species: Rattus norvegicus (Norway rat)

C/Date: 16-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 20-Aug-1999

C/Accession: S49158

R:Schwaible, W.; Petry, F.; Loos, M.  
 submitted to the EMBL Data Library, March 1993  
 A:Description: cDNA cloning and expression of the mRNA encoding the B-chain of rat C1q.  
 A:Reference number: S49158  
 A:Accession: S49158  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-253 <SCH>  
 A:Cross-references: EMBL:X71127, NID:G510191, PIDN:CAA50440.1, PID:G510192  
 C:Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal hom  
 F:121-249/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 10.3%; Score 168.5; DB 2; Length 253;  
 Best Local Similarity 29.5%; Pred. No. 1.3e-06;  
 Matches 46; Conservative 36; Mismatches 59; Indels 15; Gaps 7;

QY 2 GPTGPG-----GSSRLRS-----AFSAARTPLEGTSEMAVPEKYYVIGGDFVATGQF 52  
 DB 100 GP-PGPRGPKGSSGGYKATQKVAFSALRTVNSALPNQALIREKVTITVNDYEPKSGKF 158  
 QY 53 RCRTVGATFFSTFACKAPRKSLVWLVNRDEVQ-ALAFDEQRPRGARRAASQSMLOLD 111  
 DB 159 TCKVGLVYFTYTHASSRGNLGVNIVRGRDRMRQKVLTFCDYAQ-NTFQVITGVVLKLE 217  
 QY 112 YGDTVWLRHGHAPHALGAPGAT--FSGYLVYADAD 145  
 DB 218 QEEVHLQATD-KNSLLGVGEGANSIFGTGLFPDMD 252

RESULT 11  
 S31216  
 collagen alpha 1(X) chain precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text change 13-Aug-1999  
 C:Accession: S31216, S28807, S22215, S30127, 148299, S26397, S31830  
 R:Kong, R.Y.C.; Kwan, K.M.; Lau, E.T.; Thomas, J.T.; Boot-Handford, R.P.; Grant, M.E.; C  
 Eur. J. Biochem. 213, 99-111, 1993  
 A:Title: Intronic exon structure, alternative use of promoter and expression of the mouse  
 A:Reference number: S31216; MUID:93238750; PMID:8477738  
 A:Accession: S31216  
 A:Molecule type: DNA  
 A:Residues: 1-680 <KON>  
 A:Cross-references: EMBL:X22110, NID:G49793; PIDN:CAA97936.1; PID:G49794  
 R:Elima, K.; Berola, I.; Rosati, R.; Metsaeranta, M.; Garofalo, S.; Perelae, M.; de Cro  
 Biochem. J. 289, 247-253, 1993  
 A:Title: The mouse collagen X gene: complete nucleotide sequence, exon structure and exp  
 A:Reference number: S28807; MUID:93143676; PMID:8424763  
 A:Accession: S28807  
 A:Molecule type: DNA  
 A:Residues: 1-285, 'A', 287-680 <ELI>  
 A:Cross-references: EMBL:X67348, NID:G50480; PIDN:CAA47763.1; PID:G50481  
 R:Elima, K.; Metsaeranta, M.; Kallio, J.; Perelae, M.; Berola, I.; Garofalo, S.; de Cro  
 Biochim. Biophys. Acta 1130, 78-80, 1992  
 A:Title: Specific hybridization probes for mouse alpha-2(I) and alpha-1(X) collagen mRN  
 A:Reference number: S22215; MUID:92182017; PMID:1543751  
 A:Accession: S22215  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 385-450, 'K', 452-627 <ELA>  
 A:Cross-references: EMBL:X63013, NID:G49795; PIDN:CAA44741.1; PID:G49796  
 R:Appe, S.S.; Olsen, B.R.  
 Matrix 13, 165-179, 1993  
 A:Title: Characterization of the mouse type X collagen gene.  
 A:Reference number: S30127; MUID:93261348; PMID:8492743  
 A:Accession: S30127  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-12, 'F', 14-26, 'S', 28-247, 'L', 249-285, 'A', 287-305, 'F', 307-416, 'S', 418-499, 'L  
 R:Appe, S.S.; Seidln, M.F.; Hayashi, M.; Olsen, B.R.  
 Eur. J. Biochem. 206, 217-224, 1992  
 A:Title: Cloning of the human and mouse type X collagen genes and mapping of the mouse t  
 A:Reference number: 148299; MUID:92267014; PMID:1587271  
 A:Accession: 148299

A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 52-247, 'L', 249-285, 'A', 287-305, 'F', 307-416, 'S', 418-499, 'L', 501-566, 'C', 568, 'I  
 A:Cross-references: EMBL:X65121, NID:G50482; PIDN:CAA46237.1; PID:G667031  
 R:Summers, T.A.; Irwin, M.H.; Wayne, R.; Ballan, G.  
 J. Biol. Chem. 263, 581-587, 1988  
 A:Title: Monoclonal antibodies to type X collagen. Biochemical studies using an antibody  
 A:Reference number: S26397; MUID:88087150; PMID:2826450  
 A:Accession: S26397  
 A:Molecule type: Protein  
 A:Residues: 'SDGYFSQ', 24-26, 'KQ' <SUM>  
 C:Genetics:  
 A:Gene: Collagen-1  
 A:Map position: 10  
 A:Introns: 51/3  
 C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
 C:Keywords: coll; extracellular matrix; glycoprotein; homotrimer  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:153-679/Product: collagen alpha 1(X) chain #status predicted <MAT>  
 F:153-679/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 10.3%; Score 168.5; DB 2; Length 680;  
 Best Local Similarity 25.7%; Pred. No. 4.4e-06;  
 Matches 56; Conservative 23; Mismatches 82; Indels 57; Gaps 6;

QY 129 GARGATSGVLYVADADADAPRGPAPRPRSAFSAARTSLVSGDAPGPRHQ----- 183  
 DB 470 GNPGA--PGPAGIATKGLNGFT--GPDPGPR---GHSGERGLPGPPGPPGQAVMPD 523  
 QY 184 -----PLAFDTFVNIIGDPPD 199  
 DB 524 GPFAAGRPRLSNPVSANHGVTGMPVSATVLSKAPVAVGAPIPDELINRQCHVD 583  
 QY 200 AAGVFCRLPGAVFFSTFLCKLPRKTLSTVLMKRDVQMIYDQASRRRMOGOSVM 259  
 DB 584 PRSGIFTCKIGIYFSYHV-HVKGTHVWGLYKNGFTPM--YTYDEVSKYLDQASGSAI 641  
 QY 260 LALRGDAVWLLSHDHOCYGVSNHGKITYTFSSGLVYP 297  
 DB 642 MELTENDQVWLQLPNAESNGLYSSSEYVSSFSGLVAP 679

RESULT 12  
 CGHND  
 collagen alpha 1(X) chain precursor - human  
 N:Alternate names: procollagen alpha 1(X) chain  
 C:Species: Homo sapiens (man)  
 C:Date: 22-Nov-1993 #sequence revision 03-Nov-1995 #text change 22-Jun-1999  
 C:Accession: S26396, S30086; S15826; S18249; A43901, 151870; S21856  
 R:Reichenberger, E.; Beier, F.; LuValle, P.; Olsen, B.R.; von der Mark, K.; Bertling, W.  
 FEBS Lett. 311, 305-310, 1992  
 A:Title: Genomic organization and full-length cDNA sequence of human collagen X.  
 A:Reference number: S26396; MUID:93012005; PMID:1397333  
 A:Accession: S26396  
 A:Molecule type: DNA  
 A:Residues: 1-680 <REL>  
 A:Cross-references: EMBL:X69952, EMBL:X72578, EMBL:X72579, EMBL:X72580, GB:S47714, GB:S4;  
 R:Appe, S.S.  
 submitted to the EMBL Data Library, March 1992  
 A:Reference number: S30085  
 A:Accession: S30085  
 A:Molecule type: DNA  
 A:Residues: 'TTPFGWCVWCLL', 52-680 <APT>  
 A:Cross-references: EMBL:X65120, NID:G23129  
 A:Note: the initial difference is probably due to translation of an intronic sequence  
 R:Appe, S.; Mattei, M.G.; Olsen, B.R.  
 FEBS Lett. 282, 393-396, 1991  
 A:Title: Cloning of human alpha-1(X) collagen DNA and localization of the COL10A1 gene to  
 A:Reference number: S15826; MUID:91243838; PMID:2037056  
 A:Accession: S15826  
 A:Molecule type: DNA  
 A:Residues: 561-647, 'G', 649-666 <AP2>  
 A:Cross-references: EMBL:X58879, NID:G30013; PIDN:CAA41686.1; PID:G30014

R.Thomas, J.T.; Cresswell, C.J.; Raeh, B.; Nicolai, H.; Jones, T.; Solomon, E.; Grant, M.  
Biochem. J. 280, 617-623, 1991  
A>Title: The human collagen X gene. Complete primary translated sequence and chromosomal  
A'Reference number: S18249; NID:92109659; PMID:1764025  
A'Accession: S18249  
A'Molecule type: DNA  
A'Residuals: 1-26, 'T', 28-680 <THO>  
A'CROSS-references: EMBL:X60382; NID:930094; PIDN:CAA42933.1; PID:930095  
A'Note: the sequence from Fig. 3 is inconsistent with that from Fig. 2 in having 532-ala  
R.Reichenberger, E.; Aigner, T.; von der Mark, K.; Stoss, H.; Bertling, W.  
Dev. Biol. 148, 562-572, 1991  
A>Title: In situ hybridization studies on the expression of type X collagen in fetal hum  
A'Reference number: A43901; NID:9207285; PMID:1743401  
A'Accession: A43901  
A'Molecule type: mRNA  
A'Residuals: 547-656 <R2>  
A'CROSS-references: GB:M74050; GB:D57494; NID:9339884; PIDN:AAA61221.1; PID:9553796  
A'Note: sequence extracted from NCBI backbone (NCBI:69012, NCBI:69014)  
R.Wallis, G.A.; Raeh, B.; Sweetman, W.A.; Thomas, J.T.; Super, M.; Evans, G.; Grant, M.E.  
Am. J. Hum. Genet. 54, 169-178, 1994  
A>Title: Amino acid substitutions of conserved residues in the carboxyl-terminal domain  
pe Schmid.  
A'Reference number: I51870; NID:94136476; PMID:8304336  
A'Accession: I51870  
A>Status: translated from GB/EMBL/DBJ  
A'Molecule type: mRNA  
A'Residuals: 520-597, 'D', 599-680 <MAL>  
A'CROSS-references: GB:S68531; NID:9545180; PIDN:AAC60615.1; PID:9545181  
A'Note: mutant sequence from patient with metaphyseal chondrodysplasia type Schmid  
A'Note: a second mutant sequence with 614-Pro is also described  
C'Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
ed and subsequently O-glycosylated.  
C'Genetics:  
A'Gene: GDB:COL10A1  
A'CROSS-references: GDB:128635; OMIM:120110  
A'Map position: 6q21-6q22  
A'Introns: 52/1  
A'Note: a defect in this gene may cause Schmid metaphyseal chondrodysplasia  
C'Complex: type X collagen may be a homotrimer  
C'Function:  
A'Description: structural component of extracellular fibrous polymer specifically and to  
be important for skeletogenesis  
C'Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
C'Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer; hydroxylysine;  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-680/Product: collagen alpha 1(X) chain #status predicted <MAT>  
F:19-56/Domain: amino-terminal nonhelical #status predicted <NC2>  
F:57-519/Region: interrupted helical  
F:520-680/Domain: amino-terminal nonhelical #status predicted <NC1>  
F:553-679/Domain: complement C1q carboxyl-terminal homology <C1Q>  
F:617/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 10.2%; Score 166.5; DB 1; Length 680;  
Best Local Similarity 25.7%; Pred. No. 6.4e-06;  
Matches 56; Conservative 22; Mismatches 83; Indels 57; Gaps 6;

QY 129 GARGATGSGYLVYADADAPARPPAPRPPRSASFAARTSLVGSADAGPRHQ----- 183  
DB 470 GSGPP--PGPAGIATKGLNGPT-GRPGPPGPR---GHSGEPLGPPGPPGPGQAVME 523  
QY 184 -----PLAFTEFVNIGDDP 199  
DB 524 GRIKAGORSLSTPLVSNAGVTGMPVSAFTVILSKAYPAIGTPIPFDKILYNKQOHYD 583  
QY 200 AAGVFRCLPGAYFPFSLGKLPKRTLSVKLMKNRDEVQAMTYDDGASRRRMOQSVM 259  
DB 584 PRGTGTCQIPGIYYSYHV-HVKGTHVWVGLKNGTPTWV-YYDEYTGVLDDQASGSAI 641  
QY 260 LALRGDAVWLSHDHGIGVANSNGKRYTTFSGFLVYP 297  
DB 642 IDLTENDQVWLQIPNAESNGLYSSEYVHSSFSGLVAP 679

RESULT 13  
S23779  
collagen alpha 1(VIII) chain - mouse  
C'Species: Mus musculus (house mouse)  
C'Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
C'Accession: S23779  
R.Muragaki, Y.; Shiota, C.; Inoue, M.; Ooshima, A.; Olsen, B.R.; Ninomiya, Y.  
Eur. J. Biochem. 207, 895-902, 1992  
A>Title: Alpha-1(VIII)-collagen gene transcripts encode a short-chain collagen polypeptic  
A'Reference number: S23779; NID:922626; PMID:1499564  
A'Accession: S23779  
A'Molecule type: DNA  
A>Status: preliminary  
A'CROSS-references: EMBL:X66976; NID:950493; PIDN:CAA47387.1; PID:91359953  
C'Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
F:616-742/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 10.1%; Score 165.5; DB 1; Length 743;  
Best Local Similarity 25.8%; Pred. No. 8.7e-06;  
Matches 56; Conservative 23; Mismatches 85; Indels 53; Gaps 6;

QY 120 LKGAHYALGAGATGSGYLVYADADAPARPPAPRPPRSASFAARTSLVGS--DA 176  
DB 540 LHGPP-----GKPGA-----LGPQGGPLGPPGPPGPPAPVAMPTPSQGEYLPDM 587  
QY 177 GPG-----PRHQPLAFTEFVNIGDDPDA 200  
DB 588 GLGIDGVKTPHAYACKKGGSPAYEMPAFTALETVPPPPGAPVAFDLVYNGRONTVP 647  
QY 201 AAGVFRCLPGAYFPFSLGKLPKRTLSVKLMKNRDEVQAMTYDDGASRRRMOQSVM 260  
DB 648 QGTGTCQIPGIYYSYHV-HCKGTHVWVGLKNGTPTWV-YYDEYTGVLDDQASGSAI 705  
QY 261 ALRGDAVWLSHDHGIGVANSNGKRYTTFSGFLVYP 297  
DB 706 ILRPGDQVFLQNPFEQAGLVAGYVHSSFSGLVYP 742

RESULT 14  
S13301  
collagen alpha 1(X) chain precursor - bovine  
C'Species: Bos primigenius taurus (cattle)  
C'Date: 21-Nov-1993 #sequence\_revision 23-Feb-1996 #text\_change 13-Aug-1999  
C'Accession: S13301  
R.Thomas, J.T.; Kwan, A.P.L.; Grant, M.E.; Boot-Handford, R.P.  
Biochem. J. 273, 141-148, 1991  
A>Title: Isolation of cDNAs encoding the complete sequence of bovine type X collagen. Evl  
A'Reference number: S13301; NID:9113131; PMID:1703407  
A'Accession: S13301  
A'Molecule type: mRNA  
A'Residuals: 1-674 <THO>  
A'CROSS-references: EMBL:X53556; NID:9263; PIDN:CAA37624.1; PID:9264  
C'Genetics:  
A'Gene: COL10A1  
C'Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
C'Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-674/Product: collagen alpha 1(X) chain #status predicted <MAT>  
F:547-673/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 10.0%; Score 163.5; DB 2; Length 674;  
Best Local Similarity 26.8%; Pred. No. 1.1e-05;  
Matches 57; Conservative 18; Mismatches 87; Indels 51; Gaps 6;

QY 128 LGAPATGSGYLVYADADAPARPPAPRPPRSASFAARTSLVGSADAGPRHQ----- 183  
DB 469 VGTGPP--PGPAGIATKGLNGPT-GRPGPPR---GNAAGEPLGPPGPPGPGQVALP 522  
QY 184 -----PLAFTEFVNIGDDPDAAGV 204  
DB 523 EDFVAGORPPVSNAGVTGMPVSAFTVILSKAYPAIGTPIPFDKILYNKQOHYDPRGTI 582





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:09 ; Search time 11.2455 Seconds  
(without alignments)  
1308.910 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_17\_329  
Perfect score: 1639  
Sequence: 1 LGPTGPGSSSLRSAFSAAR.....LVYDLPAPAPGGLGASL 313

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1639	100.0	329 1	CQT4_HUMAN
2	1639	100.0	329 1	CQT4_HUMAN
3	213	13.0	243 1	CQT2_HUMAN
4	201	12.3	285 1	CQT6_HUMAN
5	200	12.2	278 1	CQT6_HUMAN
6	199	12.1	254 1	CIRF_MOUSE
7	192	11.7	244 1	APM1_HUMAN
8	192	11.7	258 1	CIRF_HUMAN
9	188	11.5	247 1	APM1_MOUSE
10	186.5	11.4	255 1	GLIC_MOUSE
11	184	11.2	215 1	HP25_TAMSI
12	182	11.1	251 1	CIOB_HUMAN
13	181	11.0	289 1	CQT7_HUMAN
14	173	10.6	744 1	CA18_RABIT
15	171	10.4	419 1	COLE_TEPMA
16	170	10.4	245 1	CIOC_HUMAN
17	168.5	10.3	253 1	CIOB_RAT
18	168.5	10.3	680 1	CA1A_MOUSE
19	167.5	10.2	674 1	CA1A_CHICK
20	167	10.2	744 1	CA18_HUMAN
21	166.5	10.2	680 1	CA1A_HUMAN
22	166.5	10.2	743 1	CA18_MOUSE
23	163.5	10.0	246 1	CQT3_HUMAN
24	163.5	10.0	674 1	CA1A_BOVIN
25	163	9.9	224 1	CERL_RAT
26	160.5	9.8	246 1	CIOC_MOUSE
27	160.5	9.8	253 1	CIOB_MOUSE
28	159	9.7	245 1	CIOA_MOUSE
29	158.5	9.7	508 1	OT01_ONCKE
30	156.5	9.5	193 1	CERB_HUMAN
31	156.5	9.5	193 1	CERB_MOUSE
32	150	9.2	196 1	HP20_TAMSI
33	150	9.2	245 1	CIOA_HUMAN

34	144.5	8.8	215 1	HP27_TAMSI	Q06577 tamias sibi
35	144	8.8	1228 1	ECM_HUMAN	Q13201 homo sapien
36	139.5	8.5	635 1	CA28_HUMAN	P25067 homo sapien
37	139	8.5	201 1	CERL_HUMAN	Q9ntu7 homo sapien
38	135.5	8.3	170 1	CA28_MOUSE	P25318 mus sapien
39	98.5	6.0	265 1	Y176_HUMAN	Q14681 homo sapien
40	96.5	5.9	476 1	LEU2_STRCO	O86534 streptomyce
41	95.5	5.8	493 1	YKAT_CAEEL	P34258 caenorhabdi
42	94	5.7	1045 1	GNB_CELFI	P26225 cellulomona
43	91	5.6	673 1	ZDH8_HUMAN	Q9ulc8 homo sapien
44	91	5.6	2505 1	CCAA_HUMAN	O00555 homo sapien
45	90.5	5.5	4289 1	TENV_HUMAN	P22105 homo sapien

## ALIGNMENTS

RESULT 1					
CQT4_HUMAN	STANDARD:	PRT:	329 AA.		
ID	CQT4_HUMAN				
AC	Q9BXJ3:				
DT	28-FEB-2003 (Rel. 41, Created)				
DT	28-FEB-2003 (Rel. 41, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Complement-c1q tumor necrosis factor-related protein 4 precursor.				
CN	C1QTNF4 OR CTRP4				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Holloway J.L., Lok S.;				
RT	"Homo sapiens complement-c1q tumor necrosis factor-related protein.";				
RL	Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.				
CC	-1- SIMILARITY: Contains 2 C1Q domains.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	-----				
DR	EMBL; AF329838; AAK17962.1; -				
DR	Genew; HGNC:14346; C1QTNF4.				
DR	InterPro; IPR001073; C1Q.				
DR	Pfam; PF00386; C1Q; 2.				
DR	SMART; SM00110; C1Q; 2.				
KW	PROSITE; PS01113; C1Q; 2.				
FT	Repeat; signal.				
FT	SIGNAL	1	16	POTENTIAL.	
FT	CHAIN	17	329	COMPLEMENT-C1Q TUMOR NECROSIS FACTOR-	
FT	DOMAIN	23	159	RELATED PROTEIN 4.	
FT	DOMAIN	170	314	C1Q 1.	
FT	DOMAIN	170	314	C1Q 2.	
FT	SEQUENCE	329 AA;	35265 MW;	331CTDBF26036915 CRC64;	
Query Match	100.0%; Score 1639; DB 1; Length 329;				
Beat local similarity	100.0%; Pred. No. 1.9e-126;				
Matches 313; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
QY	1 LGPTGPGSSSLRSAFSAARTTPLEGSEMAVTFEDKVTYVNIQGFVDVATGQFRCPVPAAY 60				
DB	17 LGPTGPGSSSLRSAFSAARTTPLEGSEMAVTFEDKVTYVNIQGFVDVATGQFRCPVPAAY 76				
QY	61 FFSPTAGARPHKSLVSVMLVNRDDEVQALAFDEQRPRGARRASQSAMQLQYGDVWRL 120				
DB	77 FFSPTAGARPHKSLVSVMLVNRDDEVQALAFDEQRPRGARRASQSAMQLQYGDVWRL 136				
QY	121 HGAPHYALGAPATFSGLVVDADADAPARGPPAPPEPRSAFSAARTSLVSGDAGP 180				

Db 137 HGAPHYALGAPGTFSGYLVAADADADAPARPPAPPEPRSAFSARTRSLVGSADPAP 196  
 QY 181 RHQPLAFDTFEPVNIIGDPPDAAGVFRCLPGAVFSEFTLGKLPKRLSYLKMKNDEVQA 240  
 Db 197 RHQPLAFDTFEPVNIIGDPPDAAGVFRCLPGAVFSEFTLGKLPKRLSYLKMKNDEVQA 256  
 QY 241 MYDDASRRREKQSSQSVMLALRGDAVWVLSHDHGCGYAGSNHGXYTFSGFLVYPDLA 300  
 Db 257 MYDDASRRREKQSSQSVMLALRGDAVWVLSHDHGCGYAGSNHGXYTFSGFLVYPDLA 316  
 QY 301 PAAPGLGASELL 313  
 Db 317 PAAPGLGASELL 329

RESULT 2  
 ID COT1 HUMAN STANDARD; PRT; 281 AA.  
 AC 09BXJ1; 096NF2; 09GZRA;  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Complement-clq tumor necrosis factor-related protein 1 precursor  
 DE (G protein coupled receptor interacting protein) (GIP).  
 GN C10TNF1 OR CTRP1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RT "GIP, a putative GPCR interacting protein."  
 RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RT "Homo sapiens complement-clq tumor necrosis factor-related protein."  
 RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RX MEDLINE=2388257; PubMed=12477932;  
 RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Schemm C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mallary S.J.,  
 RA Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gnaratne P.H.,  
 RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W.,  
 RA Villalón D.K., Muzny D.A., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Ketterman M., Madan A.C., Rodriguez Y., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shcherchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Maria M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [4]  
 RP SEQUENCE OF 83-281 FROM N.A.  
 RA Iehabashi T., Kanohori K., Yosida M., Watanabe S., Ishida S., Ono Y.,  
 RA Houta T., Hirooka S., Murakawa K., Takiguchi S., Kusano J.,  
 RA Watanabe M., Fujimori K., Tanai H., Iehida M., Yamaehita H., Chiba Y.,  
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,  
 RA Yamamoto K., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,  
 RA Kimura K., Matsumoto Y., Sekine M., Kikuchi H., Kanda K.,  
 RA Wagaetuna M., Takahashi-Pujji A., Oshima A., Sugiyama A., Kawakami B.,  
 RA Suzuki Y., Sugano S., Nagahari K., Maeno Y., Negai K., Isegai T.,

RT "NEBO human cDNA sequencing project."  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Contains 1 C1Q domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AJ272138; CAC20425.1; ALT INIT.  
 DR EMBL; AF232905; AAG44303.1; ALT INIT.  
 DR EMBL; AF329840; AAK17964.1; -  
 DR EMBL; BC021553; AAH21553.1; -  
 DR EMBL; AK055541; BAB70947.1; -  
 DR Genew; HGNC:14324; C10TNF1.  
 DR InterPro; IPR001073; C1q.  
 DR InterPro; IPR000987; Collagen.  
 DR Pfam; PF00386; C1q; 1.  
 DR Pfam; PF01391; Collagen; 1.  
 DR SMART; SMO0110; C1Q; 1.  
 DR PROSITE; PS01113; C1Q; FALSE\_NEG.  
 KW Collagen; Signal.  
 FT SIGNAL 1 25  
 FT CHAIN 26 281  
 FT  
 FT DOMAIN 99 140  
 FT DOMAIN 147 281  
 FT CONFLICT 22 22 L -> P (IN REF. 1).  
 FT CONFLICT 241 241 R -> Q (IN REF. 1).  
 SQ SEQUENCE 281 AA; 31743 MW; 49E24C8B8ACB7C CRC64;

Query Match 13.1%; Score 215; DB 1; Length 281;  
 Best Local Similarity 30.2%; Pred. No. 1.7e-10;  
 Matches 77; Conservative 26; Mismatches 86; Indels 66; Gaps 11;

QY 53 RCRVPGAVFSEFTGAPKPKSLVWLVNRNDEVQALAFDEORPPARPARAASQAMLDYD 112  
 Db 75 RCDPDTGSMYPATA--VPQINTITILKGEKDR-----GDRG-----LQGRY 113  
 QY 113 GDT--VWRLHGA--HYALGAPGATFSGYLVADADADAPARPPAPPEPRSAFSAR 167  
 Db 114 GTTGSAGARGHTGPGQXGKAPBERCKSH--YA-----AFVGR 152  
 QY 168 TRSLVGSADPGRH-----PLAFDTFEPVNIIGDPPDAAGVFRCLPGAVFSEFTLGK 222  
 Db 153 KK-----PMSNHYQTVIETFEVNLXDFHNFETGKFCYVPGLYFSINVHTW 202  
 QY 223 PKRTLSVLMKNKRDVQAMITDDGASRRREKQSSQSVMLALRGDAVWVLSHDHGCGAYS 282  
 Db 203 NQKETYLLHMKNEEVLLFAQVG--DRSINQSSQSVMLALRGDAVWVLSHDHGCGAYS 260  
 QY 283 NH--GKTYTFSGFLV 295  
 Db 261 SEELDTYTFSGFLV 275

RESULT 3  
 ID COT5 HUMAN STANDARD; PRT; 243 AA.  
 AC 09BXJ0; 09UFX4;  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Complement-clq tumor necrosis factor-related protein 5 precursor.  
 GN C10TNF5 OR CTRP5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 NC NCB1\_TaxID=9606;





DR PROSITE: PS01113; ClQ; 1.  
 KW Collagen; Signal.  
 FT SIGNAL 1 16 POTENTIAL.  
 FT CHAIN 17 258 ClQ-RELATED FACTOR.  
 FT DOMAIN 67 115 COLLAGEN-LIKE.  
 FT DOMAIN 123 258 ClQ.  
 SO SEQUENCE 258 AA; 26485 MW; F776E2D206EBF763 CRC64;  
 Query Match 12.1%; Score 199; DB 1; Length 258;  
 Best Local Similarity 27.8%; Pred. No. 3.1e-09;  
 Matches 69; Conservative 22; Mismatches 87; Indels 70; Gaps 8;

QY 94 RRPGRARRASQSMQLVYGVDMRLHGAPHYAL-----GAPGTFSGYLVYADADDA 148  
 DB 38 RGPAGARS-----DGGDVSSQSGAPPSTVGPQKPKRT----- 75  
 QY 149 PARGPAPPEPRSAFSAARTSLVSDAGPGRHP----- 184  
 DB 76 GKPGPPGPRGPRGPPVGPPEKGPCKPGRGLPGSGSGAISTATYTPRYAFYAG 135  
 QY 185 -----LAFTEFVNIGSDPDAAAGVPRCLPGAYFFSFTLGKLP---KTLGVK 230  
 DB 136 LKNPHEGYVLKFDVVNTLGNVYDAAGKFTCNIPGTFFTYHY--LMRGDGTSMWAD 193  
 QY 231 LMKRDEYQMTYDGCASRRRMOSSVWMLARGDATWLSHHDGAGVANSHEKXTTF 290  
 DB 194 LCKN-GQVRASAIADQNDYDASVYLHLADGEVFIKLDGKAGHGSNN-KYSTF 250  
 QY 291 SGFLVYPP 298  
 DB 251 SGFTYSD 258

RESULT 7  
 APML\_HUMAN  
 ID APML\_HUMAN STANDARD; PRT; 244 AA.  
 AC Q15848;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Adiponectin precursor (30 kDa adipocyte complement-related protein) (ACRP30) (Adipose most abundant gene transcript 1) (apm-1) (Gelatin-binding protein).  
 DE ACRP30 (Adipose most abundant gene transcript 1) (apm-1) (Gelatin-binding protein).  
 OS Homo sapiens (Human).  
 OS APM1 OR ACRP30 OR GBP28.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Adipose tissue;  
 RX MEDLINE=96224171; PubMed=8619847;  
 RA Maeda K., Okubo K., Shimomura I., Funahashi T., Matsuzawa Y., Matsubara K.,  
 RT "cDNA cloning and expression of a novel adipose specific collagen-like factor, apm1 (Adipose Most abundant Gene transcript 1).";  
 RL Biochem. Biophys. Res. Commun. 221:286-289(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=99196984; PubMed=10095105;  
 RA Saito K., Tobe T., Minoshima S., Asakawa S., Sumiya J., Yoda M., Nakano Y., Shimizu N., Tomita M.,  
 RT "Organization of the gene for gelatin-binding protein (GBP28).";  
 RL Gene 229:67-73(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=99333693; PubMed=10403784;  
 RA Schaeffler A., Orso E., Palitzsch K.D., Buechler C., Drobnik W., Fuest A., Schoelmerich J., Schmitz G.,  
 RT "The human apm-1, an adipocyte-specific gene linked to the family of TNFs and to genes expressed in activated T cells, is mapped to chromosome 1q21.3-q23, a susceptibility locus identified for familial combined hyperlipidemia (FCH).";

RL Biochem. Biophys. Res. Commun. 260:416-425(1999).  
 RN [4]  
 RP CHARACTERIZATION.  
 RX MEDLINE=20417747; PubMed=10961870;  
 RA Yokota T., Oritani K., Takahashi I., Ishikawa J., Matsuyama A., Ouchi N., Kihara S., Funahashi T., Tenner A.J., Tomiyama Y., Matsuzawa Y.;  
 RT "Adiponectin, a new member of the family of soluble defense collagens, negatively regulates the growth of myelomonocytic progenitors and the functions of macrophages.";  
 RL Blood 96:1723-1732(2000).  
 RN [5]  
 RP CHARACTERIZATION.  
 RX MEDLINE=20440368; PubMed=10982546;  
 RA Ouchi N., Kihara S., Arita Y., Okamoto Y., Maeda K., Kuriyama H., Hotta K., Nishida M., Takahashi M., Muraguchi M., Ohmoto Y., Nakamura T., Yamashita S., Funahashi T., Matsuzawa Y.;  
 RT "Adiponectin, an adipocyte-derived plasma protein, inhibits endothelial NF-kappaB signaling through a C/EBP-dependent pathway.";  
 RL Circulation 102:1296-1301(2000).  
 RN [6]  
 RP FUNCTION.  
 RX MEDLINE=21372498; PubMed=11479627;  
 RA Yamauchi T., Kamon J., Maki H., Terachi Y., Kubota N., Hara K., Mori Y., Ide T., Murakami K., Tsuboyama-Kasaoka N., Ezaki O., Akanuma Y., Gavrilova O., Vinson C., Reitman M.L., Kagechika H., Shudo K., Yoda M., Nakano Y., Tobe K., Nagai R., Kimura S., Tomita M., Froguel P., Kadowaki T.;  
 RT "The fat-derived hormone adiponectin reverses insulin resistance associated with both lipodystrophy and obesity.";  
 RL Nat. Med. 7:941-946(2001).  
 RN [7]  
 RP VARIANT ADIPONECTIN DEFICIENCY CYS-112.  
 RX MEDLINE=20378830; PubMed=10918532;  
 RA Takahashi M., Arita Y., Yamagata K., Matsukawa Y., Okutomi K., Horie M., Shimomura I., Hotta K., Kuriyama H., Kihara S., Nakamura T., Yamashita S., Funahashi T., Matsuzawa Y.;  
 RT "Genomic structure and mutations in adipose-specific gene, adiponectin.";  
 RL Int. J. Obes. Relat. Metab. Disord. 24:861-868(2000).  
 RN [8]  
 RP VARIANTS ARG-84; MET-117; THR-164; SER-221 AND PRO-241.  
 RX MEDLINE=21671103; PubMed=11812766;  
 RA Hara K., Boutin P., Mori Y., Tobe K., Dina C., Yasuda K., Yamauchi T., Otake S., Okada T., Eto K., Kadowaki H., Hagiura R., Akanuma Y., Yazaki Y., Nagai R., Taniyama M., Matsubara K., Yoda M., Nakano Y., Kimura S., Tomita M., Kimura S., Ito C., Froguel P., Kadowaki T.;  
 RT "Genetic variation in the gene encoding adiponectin is associated with an increased risk of type 2 diabetes in the Japanese population.";  
 RL Diabetes 51:536-540(2002).  
 CC -1- FUNCTION: IMPORTANT NEGATIVE REGULATOR IN HEMATOPOIESIS AND IMMUNE SYSTEMS; MAY BE INVOLVED IN ENDING INFLAMMATORY RESPONSES THROUGH ITS INHIBITORY FUNCTIONS. INHIBITS ENDOTHELIAL NF-KAPPAB SIGNALING THROUGH A CAMP-DEPENDENT PATHWAY. INHIBITS TNF-ALPHA-INDUCED EXPRESSION OF ENDOTHELIAL ADHESION MOLECULES. INVOLVED IN THE CONTROL OF FAT METABOLISM AND INSULIN SENSITIVITY.  
 CC -1- SUBUNIT: HOMODIGLIMER (POTENTIAL).  
 CC -1- TISSUE SPECIFICITY: Secreted.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- DISEASE: Defects in APM1 are the cause of adiponectin deficiency [MIM:605441], resulting in very low concentration of plasma adiponectin. Decreased adiponectin plasma levels are associated with obesity insulin resistance, and diabetes type 2.  
 CC -1- PHARMACEUTICAL: Adiponectin might be used in the treatment of diabetes type 2 and insulin resistance.  
 CC -1- SIMILARITY: Contains 1 collagenous domain.  
 CC -1- SIMILARITY: Contains 1 ClQ domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D45371; BAA08237.1; -  
DR EMBL; AB012165; BAA66716.1; -  
DR EMBL; AB012164; BAA66716.1; JOINED.  
DR EMBL; AJ131460; CAB52413.1; -  
DR EMBL; AJ131461; CAB52413.1; JOINED.  
DR PIR; JC4708; JC4708.  
DR MIM; 605441; -  
DR GO; 0006091; P-energy pathways; TAS.  
DR InterPro; IPRO01073; C1q.  
DR InterPro; IPRO00087; Collagen.  
DR Pfam; PF00386; C1q; 1.  
DR Pfam; PF00391; Collagen; 1.  
DR PRINTS; PRO0007; COMPLEMENTC1Q.  
DR ProDom; PD000007; C1q\_helix; 1.  
DR SMART; SM00110; C1Q; 1.  
DR PROSITE; PS01113; C1Q; 1.  
KW Hormone; Collagen; Signal; Repeat; Hydroxylation; Plasma; Polymorphism; Disease mutation; Obesity; Diabetes mellitus.  
FT SIGNAL; 1 14  
FT CHAIN; 15 244  
FT DOMAIN; 42 107  
FT DISULFID; 108 244  
FT MOD\_RES; 36 36  
FT MOD\_RES; 44 44  
FT MOD\_RES; 47 47  
FT MOD\_RES; 53 53  
FT MOD\_RES; 62 62  
FT MOD\_RES; 71 71  
FT MOD\_RES; 76 76  
FT MOD\_RES; 86 86  
FT MOD\_RES; 95 95  
FT MOD\_RES; 104 104  
FT MOD\_RES; 84 84  
FT VARIANT; 112 112  
FT VARIANT; 117 117  
FT VARIANT; 164 164  
FT VARIANT; 221 221  
FT VARIANT; 241 241  
FT SEQUENCE; 244 AA; 26414 MW; 64DBCG61204B1018 CRC64;  
Query Match 11.7%; Score 192; DB 1; Length 244;  
Best Local Similarity 31.3%; Pred. No. 1.1e-08;  
Matches 47; Conservative 30; Mismatches 55; Indels 18; Gaps 6;  
QY 5 PPPSSSLSSAASARTTPLEGTSEMAVTFEDKYVNVIGGDPVYVQFCRCRGAFFSF 64  
DB 104 PEEGAVVVSAPSVGLETVVT-IIPNPIRFTKIFVYQNHVGSCTKFCFKNIPGLVYFV 162  
QY 65 TAGKAPH-----KSLSVMLVRNRDEVALPDEQRPPARRAASOSAMQLDYGDTVM 119  
DB 163 -----HITVYWKDVVSLFK-KDKAMLFYQYQYQYQYQYQYQYQYQYQYQYQ 214  
QY 120 LMGAPH-----VALGAPGATFSGVLYVADAD 145  
DB 215 VYGEGBRNGLYVADNDNDSTFTGFLVHDTN 244  
RESULT 8  
CIRF\_HUMAN STANDARD; PRT; 258 AA.  
AC 075973;  
DT 16-OCT-2001 (Rel. 40, Created)  
FT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE C1q-related factor precursor.  
GN C1QRF OR CRF.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99097006; PubMed=9878755;  
RA Berube N.G., Swanson X.H., Bertam M.J., Kittle J.D., Didenko V.,  
RA Baeklin D.S., Smith J.R., Pereira-Smith O.M.;  
RT "Cloning and characterization of CRF, a novel C1q-related factor,  
RT expressed in areas of the brain involved in motor function.";  
RN Brain Res. Mol. Brain Res. 63:233-240(1999).  
[2]  
RP SEQUENCE FROM N.A.  
RA Rieder M.J., Carrington D.P., Chung M.-W., Lee K.L., Poel C.L., Yi Q.,  
RA Nickerson D.A.;  
RN Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
[3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Straube R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Martuna B., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshlyuki S., Carrinci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAINSTEM.  
CC -1- SIMILARITY: Contains 1 collagenous domain.  
CC -1- SIMILARITY: Contains 1 C1q domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF095154; AAC64186.1; -  
DR EMBL; AF410771; AAK95248.1; -  
DR EMBL; BC008798; AAO8798.1; -  
DR GO; 0007626; P: locomotor behavior; NAS.  
DR InterPro; IPRO01073; C1q.  
DR InterPro; IPRO00087; Collagen.  
DR Pfam; PF00386; C1q; 1.  
DR Pfam; PF00391; Collagen; 1.  
DR PRINTS; PRO0007; COMPLEMENTC1Q.  
DR SMART; SM00110; C1Q; 1.  
DR PROSITE; PS01113; C1Q; 1.  
KW Collagen; Signal.  
FT SIGNAL; 1 16  
FT CHAIN; 17 258  
FT DOMAIN; 67 115  
FT SEQUENCE; 258 AA; 26452 MW; 52C51CDF59CAE2BF CRC64;  
POTENTIAL.  
C1Q-RELATED FACTOR.  
COLLAGEN-LIKE.  
C1Q.





Polymorphism; 3d-structure.

FT SIGNAL 1 17 POTENTIAL.

FT CHAIN 18 247 ADIPONECTIN.

FT DOMAIN 45 110 COLLAGEN-LIKE.

FT DISULFID 111 247 C1Q.

FT MOD RES 39 39 INTERCHAIN (BY SIMILARITY).

FT MOD RES 47 47 HYDROXYLATION (BY SIMILARITY).

FT MOD RES 50 50 HYDROXYLATION (BY SIMILARITY).

FT MOD RES 56 56 HYDROXYLATION (BY SIMILARITY).

FT MOD RES 65 65 HYDROXYLATION (BY SIMILARITY).

FT MOD RES 79 79 HYDROXYLATION (BY SIMILARITY).

FT MOD RES 98 98 HYDROXYLATION (BY SIMILARITY).

FT MOD RES 107 107 HYDROXYLATION (BY SIMILARITY).

FT VARIANT 113 113 M -> V.

FT CONFLICT 50 50 P -> S (IN REF. 2).

FT CONFLICT 74 74 A -> S (IN REF. 2).

FT CONFLICT 117 117 A -> G (IN REF. 2).

FT CONFLICT 148 148 G -> N (IN REF. 2).

FT CONFLICT 243 243 Y -> F (IN REF. 2).

SEQUENCE 247 AA; 26841 MW; 137B687D873988C4 CRC64;

Query Match 11.5%; Score 188; DB 1; Length 247;

Best Local Similarity 33.5%; Pred. No. 2.3e-08;

Matches 53; Conservative 24; Mismatches 73; Indels 8; Gaps 5;

QY 146 ADAPRANGPAPP-----EPSSAFSAATRLVSGSDAPGRHROPLADTEFVNIGDPDA 201

DB 91 AAGP-RGPGTGPGRKGEPEAAVYSAFVSGLETRVTVNPFIRTKLFYNOQNHDS 149

QY 202 AGVFCRLGAAVFPFTLGKLPKTLGVLMKNRDEVMAMITDDGASRRRMOSSVMLA 261

DB 150 TGRFYNIIGLYFSFHT-TVMKDVKSLFK-KDAVLFITDIOYKRVNDQSSSVLLH 207

QY 262 LRGDVWLLSH-DHDYGAYSNHGKXITPFGFLVYPD 298

DB 208 LEVGQVWLVGVYGDGDHNGLYADNVNDSTFTGFLYHD 245

RESULT 10

GLIC MOUSE STANDARD; PRT; 255 AA.

AC Q9ESR4;

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Gliacolin precursor (C1q-like protein).

GN C1QL.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20428709; PubMed=10862616;

RA Koide T., Aso A., Yoshizuri T., Nagata K.;

RT "Conformational requirements of collagenous peptides for recognition by the chaperone protein HSP47."

RL J. Biol. Chem. 275:27957-27963(2000).

CC -1- TISSUE SPECIFICITY: EXPRESSED IN GLIAL CELLS.

CC -1- SIMILARITY: Contains 1 collagenous domain.

CC -1- SIMILARITY: Contains 1 C1Q domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AB044560; BAB15806.1; -

DR MGI; 2387350; C1ql.

DR GO; GO:0005515; F:protein binding activity; IPI.

InterPro; IPR001073; C1q.

DR InterPro; IPR000087; Collagen.

DR Pfam; PF00386; C1q; 1.

DR Pfam; PF01391; Collagen; 1.

DR PRINTS; PRO0007; COMPLEMENTC1Q.

DR SMART; SM00110; C1Q; 1.

DR PROSITE; PS01113; C1Q; 1.

KW Collagen; signal.

FT SIGNAL 1 20 POTENTIAL.

FT CHAIN 21 255 GLIACOLIN.

FT DOMAIN 61 111 COLLAGEN-LIKE.

FT DOMAIN 120 255 C1Q.

SEQUENCE 255 AA; 26687 MW; 529FBAF4B2191BC1 CRC64;

Query Match 11.4%; Score 186.5; DB 1; Length 255;

Best Local Similarity 26.3%; Pred. No. 3.2e-08;

Matches 57; Conservative 23; Mismatches 86; Indels 51; Gaps 6;

QY 120 LHGAPHYALGAPGATFSGVLYVADADAPAPAPPEPRSAFSAATRLVSGDAPG 179

DB 52 MSLPTPIQGPKGE-----AGRPGRAGPRGPRGPRGPRGPRGPRGPRGPRG 103

QY 180 PRHQP-----LAFDTEFVNIGDPDAAGVF 205

DB 104 PGAPGLNAGAIASATYSTVEKIAFYAGLKRQHEGYELKRDVTVLNHNYDPTTKF 163

QY 206 RRLRGATFPSTLGKLR-----KTLGVLMKNRDEVMAMITDDGASRRRMOSSVMLA 261

DB 164 TCSIPGIYFFTYHV-LNRGGDGTSMWADLCN-NQVRASAIADADQNYDYASNSVLLH 220

QY 262 LRGDVWLLSHDHDYGAYSNHGKXITPFGFLVYPD 298

DB 221 LEPGDEVYIKLDGKRAHG--GNANKYTFSGFLIYAD 255

RESULT 11

HP25 TAMSI STANDARD; PRT; 215 AA.

ID HP25 TAMSI

AC Q06576;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hibernation-associated plasma protein HP-25 precursor (Hibernator-specific blood complex, 25 kDa subunit).

OS Tamias sibiricus (Siberian chipmunk) (Asian chipmunk).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae; Ox.

OX NCBI\_TaxID=64680;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93180798; PubMed=8441393;

RA Takamatsu N., Ohba K., Kondo J., Kondo N., Shiba T.;

RT "Hibernation-associated gene regulation of plasma proteins with a collagen-like domain in mammalian hibernators."

RL Mol. Cell. Biol. 13:1516-1521(1993).

CC [2]

CC SEQUENCE OF 29-62; 84-130; 172-183; 187-192 AND 201-215.

CC TISSUE=plasma;

CC MEDLINE=92112696; PubMed=1730610;

CC Kondo N., Kondo J.;

CC "Identification of novel blood proteins specific for mammalian hibernation."

CC RT "Identification of novel blood proteins specific for mammalian hibernation."

CC RL J. Biol. Chem. 267:473-478(1992).

CC -1- FUNCTION: PLASMA PROTEINS HP-20, HP-25, HP-27 AND HP-55 FORM A 140 kDa COMPLEX VIA DISULFIDE BONDS IN THE PLASMA AND ARE HIBERNATION SPECIFIC.

CC -1- SUBCELLULAR LOCATION: Extracellular.

CC -1- TISSUE SPECIFICITY: PLASMA; SYNTHESIZED IN THE LIVER.

CC -1- DEVELOPMENTAL STAGE: THE PROTEIN COMPLEX DISAPPEARS FROM THE PLASMA AT ONSET OF HIBERNATION AND REAPPEARS AS HIBERNATION CEASES.

CC -1- SIMILARITY: Contains 1 collagenous domain.  
 CC -1- SIMILARITY: Contains 1 C1Q domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: D12975; BAA02352.1;  
 CC PIR: B48150; B48150.  
 CC InterPro: IPR001073; C1Q.  
 CC InterPro: IPR000087; Collagen.  
 CC Pfam: PF00386; C1q; 1.  
 CC Pfam: PF01391; Collagen; 1.  
 CC PRINTS: PRO0007; COMPLEMENTC1Q.  
 CC SMART: SM00110; C1q; 1.  
 CC PROSITE: PS01113; C1Q; 1.  
 CC Signal: Collagen; Glycoprotein; Plasma; Multigene family.  
 CC CHAIN 1 28  
 CC FT CHAIN 29 215  
 CC FT DOMAIN 40 81  
 CC FT DOMAIN 83 215  
 CC FT CARBOHD 167 167  
 CC SEQUENCE 215 AA; 22664 MW; AFE03206917EA530 CRC64;  
 Query Match 11.2%; Score 184; DB 1; Length 215;  
 Best local similarity 30.4%; Pred No. 4.2e-08;  
 Matches 62; Conservative 22; Mismatches 52; Indels 68; Gaps 11;  
 QY 122 GAPHYALGAPGATFSGYLVADADADAPARGPAPP-----EPRS 161  
 Db 49 GIPGP-GAPGAL-----GPPGPVGVGIRPGQPGDYEKSSRRKS 90  
 QY 162 AFSARTSLVGSDDGPGRRHOPLAFTIEFVNIIGDFDAAGVFCRLPGAYFSEFTGK 221  
 Db 91 AFAVVL-----SERPPEP-FQPIVFKALYNQEGHFMWATGEFSCVPGVYNFGFDI-R 142  
 QY 222 LPRKTLVYKMKRBEVQAMITDQASRRRMOQO-----SYMLLRDGDVWLISH 273  
 Db 143 LFQSSVKRLM--KRGIOV-----REKEQANDSTKMGVYIMLGKDKXWLSK 192  
 QY 274 DHDGYGAYSNHG-KYITFSGFLVY 296  
 Db 193 LK---GTESEKGIHIVFGYLLY 213  
 RESULT 12  
 C1QB HUMAN STANDARD; PRT; 251 AA.  
 AC P02746; 096H17;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 GN Complement C1q subcomponent, B chain precursor.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN NCB1  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE=86076906; PubMed=3000358;  
 RA Reid K.B.M.;  
 RT "Molecular cloning and characterization of the complementary DNA and  
 RT gene coding for the B-chain of subcomponent C1q of the human  
 RT complement system.";  
 RL Biochem. J. 231:729-735(1985).  
 RP SEQUENCE FROM N.A.

RC TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Ditchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stepheon M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
 RA Brownstein W.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Rana S.S., Loquellano N.A., Peters G.J., Adamson R.D., Mullaly S.J.,  
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE OF 26-133.  
 RX MEDLINE=80020137; PubMed=486087;  
 RA Reid K.B.M.;  
 RT "Complete amino acid sequences of the three collagen-like regions  
 RT present in subcomponent C1q of the first component of human  
 RT complement.";  
 RL Biochem. J. 179:367-371(1979).  
 RN [4]  
 RP SEQUENCE OF 26-193.  
 RX MEDLINE=79041552; PubMed=708376;  
 RA Reid K.B.M., Thompson E.O.P.;  
 RT "Amino acid sequence of the N-terminal 108 amino acid residues of the  
 RT B chain of subcomponent C1q of the first component of human  
 RT complement.";  
 RL Biochem. J. 173:863-868(1978).  
 RN [5]  
 RP SEQUENCE OF 134-251.  
 RX MEDLINE=82283890; PubMed=6981411;  
 RA Reid K.B.M., Gagnon J., Frampton J.;  
 RT "Completion of the amino acid sequences of the A and B chains of  
 RT subcomponent C1q of the first component of human complement.";  
 RL Biochem. J. 203:559-569(1982).  
 RN [6]  
 RP SEQUENCE OF 224-251 FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=85038855; PubMed=6208566;  
 RA Reid K.B.M., Bentley D.R., Wood K.J.;  
 RT "Cloning and characterization of the complementary DNA for the B  
 RT chain of normal human serum C1q.";  
 RL Phillos. Trans. R. Soc. Lond. B, Biol. Sci. 306:345-354(1984).  
 RN [7]  
 RP REVIEW OF C1Q DEFICIENCY.  
 RX MEDLINE=98450587; PubMed=9777412;  
 RA Peetry F.;  
 RT "Molecular basis of hereditary C1q deficiency.";  
 RL Immunobiology 199:286-294(1998).  
 CC -1- FUNCTION: C1Q ASSOCIATES WITH THE PROENZYMES C1R AND C1S TO YIELD  
 CC C1, THE FIRST COMPONENT OF THE SERUM COMPLEMENT SYSTEM. THE  
 CC COLLAGEN-LIKE REGIONS OF C1Q INTERACT WITH THE CA(2+)-DEPENDENT  
 CC C1R(2)C1S(2) PROENZYME COMPLEX, AND EFFICIENT ACTIVATION OF C1  
 CC TAKES PLACE ON INTERACTION OF THE GLOBULAR HEADS OF C1Q WITH THE  
 CC FC REGIONS OF IGG OR IGM ANTIBODY PRESENT IN IMMUNE COMPLEXES.  
 CC -1- SUBUNIT: C1 IS A CALCIUM-DEPENDENT TRIMOLECULAR COMPLEX OF C1Q, R  
 CC AND S IN THE MOLAR RATION OF 1:2:2. C1Q SUBCOMPONENT IS COMPOSED  
 CC OF NINE SUBUNITS, SIX OF WHICH ARE DISULFIDE-LINKED DIMERS OF THE  
 CC A AND B CHAINS, AND THREE OF WHICH ARE DISULFIDE-LINKED DIMERS OF  
 CC THE C CHAIN.  
 CC -1- PTM: O-linked glycans consist of Glc-Gal disaccharides bound to  
 CC the oxygen atom of post-translationally added hydroxyl groups.  
 CC -1- DISEASE: Defects in C1QB are a cause of C1q deficiency

[MIM:120570]. It is a rare genetic disorder which is associated with recurrent infections and a high prevalence of lupus erythematosus-like symptoms. It is characterized by a loss of activation of the complement classical pathway.

-1- SIMILARITY: Contains 1 collagenous domain.

-1- SIMILARITY: Contains 1 C1q domain.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL; X03084; CAA26880.1; -.  
CC EMBL; BC008983; AAH08983.1; ALT\_INIT.  
CC EMBL; M36278; AAC1692.1; -.  
CC Genew; HGNC:1242; C1QB.  
CC MIM; 120570; -.  
CC GO; GO:0005602; C:complement component C1q complex; TAS.  
CC GO; GO:0003811; F:complement activity activity; TAS.  
CC InterPro; IPR001073; C1q.  
CC InterPro; IPR000087; Collagen.  
CC Pfam; PF00386; C1q; 1.  
CC Pfam; PF01391; Collagen; 1.  
CC PRINTS; PR00007; COMPLEMENTC1Q.  
CC SMART; SM00110; C1Q; 1.  
CC PROSITE; PS01113; C1Q; 1.  
CC Complement pathway; Plasma; Hydroxylation; Glycoprotein; Collagen;  
CC Repeat; Signal; Disease mutation; Pyrolydione carboxylic acid.  
CC SIGNAL; 1 25  
CC CHAIN; 26 251  
CC DOMAIN; 29 112  
CC MOD\_RES; 113 251  
CC MOD\_RES; 26 26  
CC MOD\_RES; 29 29  
CC MOD\_RES; 33 33  
CC MOD\_RES; 36 36  
CC MOD\_RES; 39 39  
CC MOD\_RES; 42 42  
CC MOD\_RES; 44 44  
CC MOD\_RES; 51 51  
CC MOD\_RES; 54 54  
CC MOD\_RES; 57 57  
CC MOD\_RES; 57 57  
CC CARBOHYD; 57 57  
CC MOD\_RES; 60 60  
CC CARBOHYD; 60 60  
CC MOD\_RES; 63 63  
CC MOD\_RES; 75 75  
CC MOD\_RES; 81 81  
CC MOD\_RES; 84 84  
CC MOD\_RES; 90 90  
CC MOD\_RES; 96 96  
CC CARBOHYD; 96 96  
CC MOD\_RES; 99 99  
CC MOD\_RES; 102 102  
CC MOD\_RES; 105 105  
CC MOD\_RES; 108 108  
CC CARBOHYD; 108 108  
CC VARIANT; 40 40  
CC CONFLICT; 26 26  
CC CONFLICT; 83 83  
CC CONFLICT; 98 98  
CC SEQUENCE; 251 AA; 26459 MW; 78C5752E267A0BF7 CRC64;  
Query Match 11.1%; Score 182; DB 1; Length 251;  
Best Local Similarity 29.1%; Pred. No. 7, 3e-08;  
Matches 48; Conservative 32; Mismatches 61; Indels 24; Gaps 5;

149 PARGPAPPEP-----RSAPSAARTSLVSGSDAPGPRHGLADTEFEVNIQ 195  
DB 95 PKGCGAGAPGAPGPKSGSGDYKATQKIAFSATRTINV-----FLRRDQTRFDHYITMNN 148

196 GDFDAAGVFCRPLPGAYFFSFTLKLPRKTLVYLMKGRDEVOAMI-YDDGASRRRMO 254  
DB 149 NNVEPRSRKFTCKPGLYFTYTHASS--RGNLCVNLAMGRBAQKVFYFCIAYVTFQVT 206

255 SOSVWMLARRDVAWVLLSHDHGGAVSNHGRKYYTFSGFLVYPLD 239  
DB 207 TGGWVLKLEGENVFLQATDKN--SLGMEGANISFGFLFPDM 249

RESULT 13  
C1Q\_HUMAN STANDARD; PRT; 289 AA.  
ID C1Q\_HUMAN  
AC 09BXJ2;  
DT 28-FEB-2003 (Rel. 41, last sequence update)  
DT 15-SEP-2003 (Rel. 42, last annotation update)  
DE Complement-c1q tumor necrosis factor-related protein 7 precursor.  
GN C1QTNF7 OR CTRP7.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Piddington C.S., Sheppard P.O., Bishop P., Lasser G.W.;  
RT "Homo sapiens complement-c1q tumor necrosis factor-related protein.";  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
RN (2)  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Bueltow K.H., Scheeler C.F., Blat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diachenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carrinci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hilyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Heltan E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smallie D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
CC -1- SIMILARITY: Contains 1 collagenous domain.  
CC -1- SIMILARITY: Contains 1 C1q domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF329839; AAK17963.1; -.  
CC EMBL; BC022187; AAH22187.1; -.  
CC Genew; HGNC:14342; C1QTNF7.  
CC InterPro; IPR001073; C1q.  
CC InterPro; IPR000087; Collagen.  
CC Pfam; PF00386; C1q; 1.  
CC Pfam; PF01391; Collagen; 2.  
CC PRINTS; PR00007; COMPLEMENTC1Q.  
CC SMART; SM00110; C1Q; 1.  
CC PROSITE; PS01113; C1Q; 1.

KW Collagen; Signal. 16 POTENTIAL.  
FT SIGNAL 17 COMPLEMENT-C1Q TUMOR NECROSIS FACTOR-  
FT CHAIN 17 RELATED PROTEIN 7.  
FT DOMAIN 38 139 COLLAGEN-LIKE.  
FT DOMAIN 141 276 C1Q.  
SQ SEQUENCE 289 AA; 30683 MW; A61609PFE6D26946 CRC64;  
Query Match 11.0%; Score 181; DB 1; Length 289;  
Best Local Similarity 33.7%; Pred. No. 1e-07;  
Matches 57; Conservative 21; Mismatches 59; Indels 32; Gaps 8;  
OY LGPTGPGP-----GSSRLRSFSAART--PLGTSMTATPKYVN 40  
DB 118 IGP-PGPGDRGEQDPLPGVCGSYLKSASFVGITTSYER--RLPIIFNKVLFN 173  
OY 41 ICGDFVATGQFRCPVCAVFPFSTAGKAPHSLSVLRNDEVQALAFDEORRPGARR 100  
DB 174 EGEHNPAATGKRCIAFPGIYFSYDITLA-NGLIALGLVHN-GQYRITFFD--ANTGNHD 229  
OY 101 AASGAMQLDYGDTVMRLHGAPHYAL---GAPGATFSGYLVYADAD 145  
DB 230 VASGSTVILYLPEDDEVMLEIFFTDONGLFSDPGMADSLFSGFLYVDT 278  
RESULT 14  
ID CA18 RABIT STANDARD; PRT; 744 AA.  
AC P14282;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Collagen alpha 1(VIII) chain precursor (Endothelial collagen).  
GN COL8A1.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
CC NCBI\_TaxId=9986;  
OK NCBI\_TaxId=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89380199; Pubmed=2476437;  
RA Yamaguchi N., Benya P.D., van der Rest M., Ninomiya Y.;  
RT "The cloning and sequencing of alpha 1(VIII) collagen cDNAs  
demonstrate that type VIII collagen is a short chain collagen and  
contains triple-helical and carboxyl-terminal non-triple-helical  
RT domains similar to those of type X collagen.";  
RL J. Biol. Chem. 264:16022-16029(1989).  
CC -I- FUNCTION: MAJOR COMPONENT OF THE DESCENDING MEMBRANE (BASEMENT  
MEMBRANE) OF CORNEAL ENDOTHELIAL CELLS.  
CC -I- SUBUNIT: MAY FORM HOMOTRIMERS, OR HETEROTRIMERS IN ASSOCIATION  
CC WITH ALPHA 2(VIII) TYPE COLLAGENS.  
CC -I- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -I- MISCELLANEOUS: 4 consecutive G-P-P tripeptides are present at the  
CC C-terminus of the triple-helical region. These may provide high  
CC thermal stability of this region.  
CC -I- SIMILARITY: STRONG, TO ALPHA 2 TYPES VIII AND X COLLAGENS.  
CC -I- SIMILARITY: Contains 1 C1Q domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; J05042; AAA31204.1;  
DR PIR; A34246; A34246.  
DR InterPro; IPR001073; C1q.  
DR InterPro; IPR000087; Collagen.  
DR Pfam; PF00386; C1q; 1.  
DR Pfam; PF01391; Collagen; 8.

DR PRINTS; PR00007; COMPLEMENTC1Q.  
DR SMART; SMO0110; C1Q; 1.  
DR PROSITE; PS01113; C1Q; 1.  
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Glycoprotein; Cell adhesion; Collagen; Signal.  
FT SIGNAL 1 20  
FT CHAIN 21 744 COLLAGEN ALPHA 1(VIII) CHAIN.  
FT DOMAIN 29 117 NONHELICAL REGION (NC2).  
FT DOMAIN 118 571 TRIPLE-HELICAL REGION.  
FT DOMAIN 572 744 NONHELICAL REGION (NC1).  
FT DOMAIN 609 744 C1Q.  
SQ SEQUENCE 744 AA; 73358 MW; 2A8CE1FE8274E99 CRC64;  
Query Match 10.6%; Score 173; DB 1; Length 744;  
Best Local Similarity 25.3%; Pred. No. 1.4e-06;  
Matches 56; Conservative 27; Mismatches 78; Indels 60; Gaps 8;  
OY 120 LHGAPHYALGAPATFGSYLVYADADADAPA-RGPPAPPEPRSS----- 161  
DB 540 LHGPP-----GKPGA-----LGPGQGPRLPGPPPGPPGPPAVMPPTPAPQGEVLPDMG 589  
OY 162 -----AFSARTSLVGSADAGC-----PRHPLAFDTFEVNIIG 196  
DB 590 LGIDGVKTPHAYAAKK-----GKNGGPAYEMPAPFAETLAPFPVGAPIKFDRLVNGRQ 644  
OY 197 DFDAAAGYFRCLGAYFSTTLGKPRKTLISVLMKRDVQAMIVDQASRRRMSQ 256  
DB 645 NYPQGTGIFTEGVGVYFAFHV-HCKGNVWVALFKNPEVM-YTYDEYKKGFLDQASG 702  
OY 257 SVMALRRGDAVWMLSHDHDGAYSNHGKTYTFSGFVLP 297  
DB 703 SAVLLLRGDRVFLQMPSEQAAGLYAGGYVHSSFSGYLLP 743  
RESULT 15  
ID COLE LEPPA STANDARD; PRT; 419 AA.  
AC P98085; Q91080;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Inner ear-specific collagen precursor (Saccular collagen).  
OS Lepomis macrochirus (Bluegill).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
CC Acanthomorphi; Acanthopterygii; Perciformes; Percoidae;  
CC Centrarchidae; Lepomis.  
CC NCBI\_TaxId=11106;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95167486; Pubmed=7863331;  
RA Davis J.G., Oberholzer J.C., Burns F.R., Greene M.I.;  
RT "Molecular cloning and characterization of an inner ear-specific  
RT structural protein.";  
RL Science 267:1031-1034(1995).  
RN [2]  
RP CONCEPTUAL TRANSLATION.  
RA Gibson T.;  
RL Submitted (MAR-1995) to the SWISS-PROT data bank.  
CC -I- FUNCTION: FORMS A MICROSTRUCTURAL MATRIX WITHIN THE OTOLITHIC  
CC MEMBRANE (PROBABLY).  
CC -I- TISSUE SPECIFICITY: SPECIALIZED SECRETORY SUPPORTING CELLS AT THE  
CC OUTER PERIMETER OF THE SACCULAR EPITHELIUM.  
CC -I- SIMILARITY: Contains 1 C1Q domain.  
CC -I- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT HAD TO BE  
CC INTRODUCED FOR POSITIONS 391-419 SO AS TO MAXIMIZE THE SIMILARITY  
CC WITH OTHER SHORT-CHAIN COLLAGENS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:10 ; Search time 43.1078 Seconds  
(without alignments)  
1873.686 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_17\_329

Perfect score: 1639  
Sequence: 1 LGPTPGSSSELRSAFSAAR.....LVYPLAPAPPGGLGASLL 313

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP rvirus:\*  
16: SP bacterioplasmid:\*  
17: SP archaeoplasmid:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1631	99.5	329	4	Q8IV25
2	1514	92.4	326	11	Q8R066
3	1003.5	61.2	205	11	Q9D0M2
4	997.5	60.9	205	11	Q9D0C6
5	230	14.0	158	4	Q9H677
6	230	14.0	1077	4	Q8TE71
7	226	13.8	158	11	Q8K110
8	220.5	13.5	182	11	Q8R1P2
9	220.5	13.5	281	11	Q9OXF7
10	213	13.0	243	4	Q8W6P2
11	205	12.5	243	11	Q8R002
12	202	12.3	243	11	Q8K479
13	200	12.2	294	11	Q9D8U4
14	193	11.8	243	6	Q95JD7
15	189	11.5	247	11	Q8BRW2
16	186	11.3	287	11	Q8CFR0

17	185	11.3	240	6	Q95M04	Q95M04 bos taurus
18	181.5	11.1	244	11	Q8K3R4	Q8K3R4 ratu
19	180.5	11.0	264	11	Q8BKRO	Q8BKRO mus muscu
20	179.5	11.0	120	11	Q8R1Z2	Q8R1Z2 mus muscu
21	177.5	10.8	312	11	Q8CHX9	Q8CHX9 mus muscu
22	168.5	10.3	289	11	Q8BVD7	Q8BVD7 mus muscu
23	167.5	10.2	744	11	Q921S8	Q921S8 mus muscu
24	167.5	10.2	744	11	Q8BGJ6	Q8BGJ6 mus muscu
25	166	10.1	675	6	Q9N178	Q9N178 sus scrofa
26	165.5	10.1	194	6	Q95J95	Q95J95 canis fami
27	164	10.0	295	11	Q9Z1K4	Q9Z1K4 ratu
28	163	9.9	224	4	Q8IUK8	Q8IUK8 homo sapien
29	163	9.9	224	11	Q8BGU2	Q8BGU2 mus muscu
30	162.5	9.9	246	11	Q9E630	Q9E630 mus muscu
31	161	9.8	173	6	Q62789	Q62789 sus scrofa
32	159	9.7	245	11	Q9DCM6	Q9DCM6 mus muscu
33	155	9.5	196	11	Q920N0	Q920N0 camias sibi
34	152	9.3	197	11	Q9JHG0	Q9JHG0 mus muscu
35	147.5	9.0	333	4	Q8IUD4	Q8IUD4 homo sapien
36	145.5	8.9	705	4	Q8TEJ5	Q8TEJ5 homo sapien
37	141.5	8.6	1017	11	Q99K41	Q99K41 mus muscu
38	140	8.5	198	11	Q8BWF0	Q8BWF0 mus muscu
39	139	8.5	198	11	Q8BWE9	Q8BWE9 mus muscu
40	135	8.2	195	11	Q8BZ53	Q8BZ53 mytilus edu
41	134.5	8.2	213	5	P83425	P83425 mytilus edu
42	131.5	8.0	347	4	Q961H6	Q961H6 homo sapien
43	131.5	8.0	583	4	Q96G58	Q96G58 homo sapien
44	131.5	8.0	992	4	Q9UG76	Q9UG76 homo sapien
45	131.5	8.0	1016	4	Q9Y6C2	Q9Y6C2 homo sapien

#### ALIGNMENTS

RESULT 1  
ID Q8IV25 PRELIMINARY; PRT; 329 AA.  
AC Q8IV25;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Similar to C1q and tumor necrosis factor related protein 4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Butcheria; Primates; Carnivora; Homiidae; Homo.  
OC NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Struhsberg R.;  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC035628; AAH35628.1;  
SQ SEQUENCE 329 AA; 35256 MW; 16064DA8182A6732 CRC64;

Query Match	Score	Length	DB 4	Length 329
Best Local Similarity	99.7%	Pred. No. 5e-132;	Mismatches 0;	Indels 0;
Matches 312;	Conservative	0;	Mismatches 1;	Gaps 0;
QY	1	LGPTPGSSSELRSAFSAARTTPLEGISEMAVTPDKVYVNIIGDQDVATGQRCRVPAY	60	
DB	17	LGPTPGSSSELRSAFSAARTTPLEGISEMAVTPDKVYVNIIGDQDVATGQRCRVPAY	76	
QY	61	FFSFTAGAPKHSLSVMLVRNDEVQALAFDQRPRGARRASOSAMQLDYGDVWRL	120	
DB	77	FFSFTAGAPKHSLSVMLVRNDEVQALAFDQRPRGARRASOSAMQLDYGDVWRL	136	
QY	121	HGAPHYALGAPGATPSGYLVYADADAPAGPPAPPPRSFSAARTSLVSGDAGCP	180	
DB	137	HGAPHYALGAPGATPSGYLVYADADADAPAGPPAPPPRSFSAARTSLVSGDAGCP	196	
QY	181	RHOPLAFDTEFVNIGDQDAAGVFCRLPGAYFFSFTLGKLPKTLVSKLMKNDEVQA	240	
DB	197	RHOPLAFDTEFVNIGDQDAAGVFCRLPGAYFFSFTLGKLPKTLVSKLMKNDEVQA	256	

QY 241 MYDDGASRRRQSSQSVMLALRRGDVWLLSHDHGCGAYSNHGXITPFGSLVYPDLA 300  
 DB 257 MYDDGASRRRQSSQSVMLALRRGDVWLLSHDHGCGAYSNHGXITPFGSLVYPDLA 316  
 QY 301 PAAPGLGASELL 313  
 DB 317 PAAPGLGASELL 329

## RESULT 2

Q99066 PRELIMINARY; PRT; 326 AA.

AC 099066; 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Similar to C1q and tumor necrosis factor related protein 4.  
 GN 0710001E10R1K.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Straubeberg R.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC027315; AAH27315.1; -  
 DR MGD; MGI:1914695; 0710001E10R1K.  
 DR InterPro; IPR001073; C1q.  
 DR Pfam; PF00386; C1q; 2.  
 DR SMART; SM00110; C1Q; 2.  
 DR PROSITE; PS01113; C1Q; 2.  
 DR SEQUENCE 326 AA; 35057 MW; 72339172B7B1051A CRC64;

Query Match 92.4%; Score 1514; DB 11; Length 326;  
 Best Local Similarity 94.6%; Pred. No. 5.6e-122;  
 Matches 295; Conservative 2; Mismatches 13; Indels 2; Gaps 2;

QY 1 LGPTPGSSSELRSASFAARTTPELGTSMAVTFDKVYVNIIGDFDVAATGRCRCRVGAY 60  
 DB 17 LGPA-GPGSSSELRSASFAARTTPELGTSMAVTFDKVYVNIIGDFDVAATGRCRCRVGAY 75  
 QY 61 FFSFTGKAPKHSLSVLYVRNDEVQALAFDEQRRCGARRASQSMQLDLDGDTWMLRL 120  
 DB 76 FFSFTGKAPKHSLSVLYVRNDEVQALAFDEQRRCGARRASQSMQLDLDGDTWMLRL 135  
 QY 121 HGAHYALGAPGATFSGLYVYADADADAPARGPAPPEPRSAFSAARTSLVGSDAAGP 180  
 DB 136 HGAHYALGAPGATFSGLYVYADADADAPARGPAPPEPRSAFSAARTSLVGSDAAGP 194  
 QY 181 RHQPLAFDTEFVNIIGDFDVAAGVFRCLPGAYFFSFTLGKLPKTLVYKMKNRDEVQ 240  
 DB 195 RHQPLAFDTEFVNIIGDFDVAAGVFRCLPGAYFFSFTLGKLPKTLVYKMKNRDEVQ 254  
 QY 241 MYDDGASRRRQSSQSVMLALRRGDVWLLSHDHGCGAYSNHGXITPFGSLVYPDLA 300  
 DB 255 MYDDGASRRRQSSQSVMLALRRGDVWLLSHDHGCGAYSNHGXITPFGSLVYPDLA 314  
 QY 301 PAAPGLGASEL 312  
 DB 315 PAAPGLGASEL 326

## RESULT 3

Q99066 PRELIMINARY; PRT; 205 AA.

AC 099066; 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE 0710001E10R1K protein.  
 GN 0710001E10R1K.

OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Embryo;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishi Y.,  
 RA Atakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischnann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Stubbli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okita T., Furuno M., Aono H., Baladrelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Guerinich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzaletti J., Mombert P.,  
 RA Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Williams L.,  
 RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL; AK04340; BAB23268.1; -  
 DR MGD; MGI:1914695; 0710001E10R1K.  
 DR InterPro; IPR001073; C1q.  
 DR Pfam; PF00386; C1q; 1.  
 DR PRINTS; PR00007; COMPLEMENTC1Q.  
 DR SMART; SM00110; C1Q; 1.  
 DR PROSITE; PS01113; C1Q; 1.  
 DR SEQUENCE 205 AA; 22190 MW; B9B237793C54786D CRC64;

Query Match 61.2%; Score 1003.5; DB 11; Length 205;  
 Best Local Similarity 93.7%; Pred. No. 2.2e-78;  
 Matches 193; Conservative 1; Mismatches 11; Indels 1; Gaps 1;

QY 107 MLQDYGDTWMLRHLGAPHYALGAPATFSGLYVYADADADAPARGPAPPEPRSAFSA 166  
 DB 1 MLQDYGDTWMLRHLGAPHYALGAPATFSGLYVYADADADAPARGPAPPEPRSAFSA 59  
 QY 167 RTRSLVGSDAAGPQHPLAPDTEFVNIIGDFDVAAGVFRCLPGAYFFSFTLGKLPKRT 226  
 DB 60 RTRSLVGSDAAGPQHPLAPDTEFVNIIGDFDVAAGVFRCLPGAYFFSFTLGKLPKRT 119  
 QY 227 LSVKMKNRDEVQAMTYDDGASRRRQSSQSVMLALRRGDVWLLSHDHGCGAYSNHGX 286  
 DB 120 LSVKMKNRDEVQAMTYDDGASRRRQSSQSVMLALRRGDVWLLSHDHGCGAYSNHGX 179  
 QY 287 YTFSGFLVYPDLAPAPGLGASEL 312  
 DB 180 YTFSGFLVYPDLAPAPGLGASEL 205

## RESULT 4

Q99066 PRELIMINARY; PRT; 205 AA.

AC 099066; 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE 0710001E10R1K protein.  
 GN 0710001E10R1K.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.



RC STRAIN=C57BL/6J; TISSUE=Brain;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Araiwa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide T., Pesole G., Quackenbush J.,  
 RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,  
 RA Norone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Saesaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyszynski B., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection";  
 RL Nature 409:685-690(2001).  
 DR EMBL, AK002948; BAB22473.1;  
 DR MGI:1914695; 0710001E10Rik.  
 DR InterPro: IPR001073; Clq.  
 DR Pfam: PF00386; Clq; 1.  
 DR PRINTS; PR00007; COMPLEMENTC1Q.  
 DR SMART; SM00110; Clq; 1.  
 DR PROSITE; PS01113; Clq; 1.  
 SQ SEQUENCE 205 AA; 22215 MW; 56AD37793C437300 CRC64;

Query Match 60.9%; Score 997.5; DB 11; Length 205;  
 Best Local Similarity 93.2%; Pred. No. 7.4e-78;  
 Matches 192; Conservative 1; Mismatches 12; Indels 1; Gaps 1;

QY 107 MQLQYGPVWMLRLGAPHYALGAPGATSGYLVYADADADAPAPPPAPPPSAFSA 166  
 DB 1 MQLQYGPVWMLRLGAPHYALGAPGATSGYLVYADADADAPAPPPAPPPSAFSA 59  
 QY 167 RTRSLVSGDAGPGRPHQPLAFDTEFVNIIGDPAAGVRCRLPGAFFSFTLGKLPKRT 226  
 DB 60 RTRSLVSGDAGPGRPHQPLAFDTEFVNIIGDPAAGVRCRLPGAFFSFTLGKLPKRT 119  
 QY 227 LSVKLMKRNDEVQAMITYDDGASRRRMOQSVMALRRGDVAWLLSHDHGCGATSNHGX 286  
 DB 120 LSVKLMKRNDEVQAMITYDDGASRRRMOQSVMALRRGDVAWLLSHDHGCGATSNHGX 179  
 QY 287 YTFSGFLVYPPDLAPAPPGGLGASEL 312  
 DB 180 YTFSGFLVYPPDLAPAPPGGLGASEL 205

RESULT 5  
 09H667 PRELIMINARY; PRT; 158 AA.  
 ID 09H667;  
 AC 09H667;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE Hypothetical protein FLJ22569.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NCBI\_Taxid=9606;  
 RX SEQUENCE FROM N.A.  
 RC Tissue=Small intestine;  
 RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,  
 RA Suzuki Y., Ohtsuka T., Sugano S.,  
 RA Nakamura Y., Isonaga T., Sugano S.,  
 RT "NEO human cDNA sequencing project";  
 RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC Tissue=Uterus;  
 RA Strauberg R.;  
 RL Submitted (May-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AK026222; BAB15398.1;  
 DR EMBL; BC007520; AAH07520.1;  
 DR InterPro: IPR001073; Clq.  
 DR Pfam: PF00386; Clq; 1.  
 DR PRINTS; PR00007; COMPLEMENTC1Q.  
 DR SMART; SM00110; Clq; 1.  
 DR PROSITE; PS01113; Clq; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 158 AA; 17625 MW; 47DB10EDD6DC9760 CRC64;

Query Match 14.0%; Score 230; DB 4; Length 158;  
 Best Local Similarity 38.4%; Pred. No. 4.7e-12;  
 Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 147 DAPARGP-----PAPPPSAFSAARTSLVSGDAGPGRPHQPLAFDTEFVNIIG 196  
 DB 5 DVPYTNPAATILPVAHYVLPPOOMRVAFSAART-----SNLAPGTLDDPVIYVDDLNNIG 59  
 QY 197 DFDAAAGVRCRLPGAFFSFTLGKLP-RKTLVSKLMKRNDEVQAMITYDDGASRRRMOQS 255  
 DB 60 TFDLQGRFNCVPVNGTYFIFHMLKLVNVPVLYNLMKNEEVLVSAYANDGAP-DHETAS 118  
 QY 256 QSVMLALRRGDVAWLLSHDHGCGA-YSNHGKITYTFSGFLVYPP 298  
 DB 119 NHAILQIFQGDQIWLRLH---RGAIYGSWSKYSTFSGYLLYOD 158

Query Match 14.0%; Score 230; DB 4; Length 1077;  
 Best Local Similarity 38.4%; Pred. No. 5.9e-11;  
 Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 147 DAPARGP-----PAPPPSAFSAARTSLVSGDAGPGRPHQPLAFDTEFVNIIG 196  
 DB 924 DVPYTNPAATILPVAHYVLPPOOMRVAFSAART-----SNLAPGTLDDPVIYVDDLNNIG 978  
 QY 197 DFDAAAGVRCRLPGAFFSFTLGKLP-RKTLVSKLMKRNDEVQAMITYDDGASRRRMOQS 255  
 DB 979 TFDLQGRFNCVPVNGTYFIFHMLKLVNVPVLYNLMKNEEVLVSAYANDGAP-DHETAS 1037  
 QY 256 QSVMLALRRGDVAWLLSHDHGCGA-YSNHGKITYTFSGFLVYPP 298  
 DB 1038 NHAILQIFQGDQIWLRLH---RGAIYGSWSKYSTFSGYLLYOD 1077

RESULT 7

```

08K110
ID 08K110 PRELIMINARY; PRT; 158 AA.
AC 08K110;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Similar to hypothetical protein FLJ22569.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA Strauberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027523; AAH27523.1; -
DR InterPro; IPR001073; Clq.
DR Pfam; PF00386; Clq; 1.
DR PRINTS; PRO0007; COMPLEMENTC1Q.
DR SMART; SM00110; Clq; 1.
DR PROSITE; PS01113; Clq; 1.
KW Hypothetical protein.
SQ SEQUENCE 158 AA; 17533 MW; 86E9321C99225FCB CRC64;

Query Match 13.8%; Score 226; DB 11; Length 158;
Best Local Similarity 39.1%; Pred. No. 1e-11;
Matches 63; Conservative 18; Mismatches 68; Indels 12; Gaps 5;

QY 140 VYADADADPARGPAPPEPRSAFSAARTSLVGSAGCPHROPPLAFTPEFVNIQGD 199
DB 8 VTSPPAAILPVIHYIPQOMRVAFSAAKT-----SLATGTLDPQVFLLNLNIGET 62
QY 200 AAAGVRCRLPGAYFFSFTLGLP-RKTLISVKLMKXNDEVMYIDGASRRRMOQSGV 258
DB 63 LQLGRENCPVNGTYVIFHMLKLAIVVPLVYVNLKKEEVLVSAYANDAP-DHETASNA 121
QY 259 MLALRGDAVWLSSHHDGCGA-YSNHGKTYTFSGGLVYPPD 298
DB 122 VIQLQGDQIMLRH---RGAIYGSWMKYSTFGYLYOD 158

RESULT 8
Q8R1P2 PRELIMINARY; PRT; 182 AA.
ID Q8R1P2;
AC Q8R1P2;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Similar to RIKEN CDNA 1600017K21 gene (Fragment).
GN 1600017K21R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strauberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC023468; AAH23468.1; -
DR MGI; MGI:1919254; 1600017K21R1K.
DR InterPro; IPR001073; Clq.
DR Pfam; PF00386; Clq; 1.
DR SMART; SM00110; Clq; 1.
FT NON TER
SQ SEQUENCE 182 AA; 20863 MW; 559C73DB9517882F CRC64;

Query Match 13.5%; Score 220.5; DB 11; Length 182;
Best Local Similarity 39.2%; Pred. No. 3.7e-11;
Matches 56; Conservative 20; Mismatches 58; Indels 9; Gaps 4;

```

```

QY 161 SAFSAARTSLVGSAGCPHROPPLAFTPEFVNIQGDPAAGVRCRLPGAYFFSFTLG 220
DB 47 AAFSVGRKALHSND-----YFQPVFDFEFVNLTKHNMGTGKCYCVPGIYFSLNVH 101
QY 221 KLPRRTLSVKLMKNDEVMYIDGASRRRMOQSGVYMLALRGDAVW--LLSHDDGY 278
DB 102 TWNQKETLTHMKNEEV-VILYAO-VSDRSIMQSGSLMELREDEVMVRLFKGEREVA 159
QY 279 GAYSNHGKTYTFSGGLVYPPDLAP 301
DB 160 IFSDEFDYITFGYLVKVPASEP 182

RESULT 9
Q9QXP7 PRELIMINARY; PRT; 281 AA.
ID Q9QXP7;
AC Q9QXP7;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative secreted protein ZS1G37 (1600017K21R1K protein).
DE ZS1G37 OR 1600017K21R1K.
GN ZS1G37 OR 1600017K21R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Shepard P., Delaher T., Grant F., Chen L., Haldeman B., McKnight G.,
RA Whitmore T., O'Hara P.;
RT "Mus musculus putative secreted protein."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Placenta;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kanukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsumo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baladrelli R., Barsh G.,
RA Blake J., Boileau D., Bojunga N., Carninci P., de Bonaio M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombereite P.,
RA Norone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlschki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AF192499; AAF06664.1; -
DR EMBL; AK005484; BAB24070.1; -
DR MGI; MGI:1919254; 1600017K21R1K.
DR InterPro; IPR001073; Clq.
DR SMART; SM00110; Clq; 1.
DR Pfam; PF00386; Clq; 1.
DR Pfam; PF01391; Collagen.1.
DR PRINTS; PRO0007; COMPLEMENTC1Q.
DR SMART; SM00110; Clq; 1.
SQ SEQUENCE 281 AA; 32009 MW; C9816216DB6419E2 CRC64;

Query Match 13.5%; Score 220.5; DB 11; Length 281;
Best Local Similarity 39.2%; Pred. No. 6.5e-11;
Matches 56; Conservative 20; Mismatches 58; Indels 9; Gaps 4;

```

```

QY 161 SAFSAARTSLVGSAGCPHROPPLAFTPEFVNIQGDPAAGVRCRLPGAYFFSFTLG 220
DB 47 AAFSVGRKALHSND-----YFQPVFDFEFVNLTKHNMGTGKCYCVPGIYFSLNVH 101
QY 221 KLPRRTLSVKLMKNDEVMYIDGASRRRMOQSGVYMLALRGDAVW--LLSHDDGY 278
DB 102 TWNQKETLTHMKNEEV-VILYAO-VSDRSIMQSGSLMELREDEVMVRLFKGEREVA 159
QY 279 GAYSNHGKTYTFSGGLVYPPDLAP 301
DB 160 IFSDEFDYITFGYLVKVPASEP 182

```

DB 146 AAFSVGRKKAHSND-----YFQPVVFDTREVNLYKHFMFTGKFCYCVYPIGFFSLNHV 200  
 QY 221 KLPRKTLVSKLKNRDEVOAMTYDDGASRRREMOSQVWLALRGGDAV--LLSHDHGY 278  
 DB 201 TNNQKETYLIHINKNEBEV-VIIYAO-VSDRSIMOSQSLMMLREDEVDVRLFKERENA 258  
 QY 279 GAYSNHGKTYTFSGFLVYDPLAP 301  
 DB 259 IFSDEPDTITFSGYLVKPAEP 281

## RESULT 10

Q8N6P2 PRELIMINARY; PRT; 243 AA.

AC 08N6P2;  
 DT 01-OCT-2002 (TREMBLrel. 22, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE C1q and tumor necrosis factor related protein 5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Strausberg R.;  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC029485; AAH29485.1;  
 DR InterPro; IPR001073; C1q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1q; 1.  
 DR Pfam; PF01391; Collagen; 1.  
 DR PRINTS; PR00007; COMPLEMENTC1Q.  
 DR SMART; SM00110; C1Q; 1.  
 SQ SEQUENCE 243 AA; 25326 MW; 6D9306A0EB21B44A CRC64;

Query Match 13.0%; Score 213; DB 4; Length 243;

Best Local Similarity 34.9%; Pred. No. 2.4e-10;  
 Matches 66; Conservative 18; Mismatches 83; Indels 22; Gaps 6;

QY 122 GAPHYALGAPGATFSG---YLVDADAD---ADAPARGPAPP-----EPGSAFSAATR 169  
 DB 57 GAF-----GAPGKGGCGPGLPGPRGDPGPRGAGAGTGAAGCSVPPKSAFSAKSE 112  
 QY 170 SLVSGDAGGPRHOPLAFTTEFVNIGDPPDAAGVRCRLPGAFFSFTLGKLPKRTLSV 229  
 DB 113 SRV-----PPSDAPLPFDRVLVNEGHDVATGKTCQVPGVYFA-VHATVYASLQF 166  
 QY 230 KLKRNDEVOAMTYDDGASRRREMOSQVWLALRGGDAVWLISHDHGYGAYSNHGKTYT 289  
 DB 167 DLVKGESIASISFFQFGWPKPASLSGAMVRLPEPDQVWGVGVGYIGIYASIKTST 226  
 QY 290 FSGFLVYPD 298  
 DB 227 FSGFLVYSD 235

## RESULT 11

Q8R002 PRELIMINARY; PRT; 243 AA.

AC 08R002;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Similar to DKFZPS6B0621 protein (Hypothetical 25.4 kDa protein).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Strausberg R.;  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC023068; AAH23068.1;  
 DR EMBL; BC025174; AAH25174.1;  
 DR MGD; MGI:238598; C1qntf5.  
 DR InterPro; IPR001073; C1q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1q; 1.  
 DR Pfam; PF01391; Collagen; 1.  
 DR SMART; SM00110; C1Q; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 243 AA; 25420 MW; 498129CD051DB97B CRC64;

Query Match 12.5%; Score 205; DB 11; Length 243;  
 Best Local Similarity 32.8%; Pred. No. 1.2e-09;  
 Matches 64; Conservative 16; Mismatches 81; Indels 34; Gaps 6;

QY 122 GAPHYALGAPGATFSGYLVVDADADAPA-RGPAP-----PPRSAF 163  
 DB 57 GAF-----GAPGKGGCG-----GRPLPGPRGDPGPRGAGPMGAICPAGCSVPPPSAF 106  
 QY 164 SAARTSLVSGDAGGPRHOPLAFTTEFVNIGDPPDAAGVRCRLPGAFFSFTLGKLP 223  
 DB 107 SAKRESRV-----PPPADTLPFDRVLVNEGHDVATGKTCQVPGVYFA-VHATVY 160  
 QY 224 RKLTVSKLKNRDEVOAMTYDDGASRRREMOSQVWLALRGGDAVWLISHDHGYGAYSN 283  
 DB 161 RASLQFDLVKNGQSIASFFQFGWPKPASLSGAMVRLPEPDQVWGVGVGYIGIYAS 220  
 QY 284 HGKTYTFSGFLVYPD 298  
 DB 221 IKTSTFSGFLVYSD 235

## RESULT 12

Q8K479 PRELIMINARY; PRT; 243 AA.

AC 08K479;  
 DT 01-OCT-2002 (TREMBLrel. 22, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Complement-c1q tumor necrosis factor-related protein.  
 GN C1QTNF5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J;  
 RX MEDLINE=22135657; PubMed=12140190;  
 RA Kameya S., Hawes N.L., Chang B., Heckenlively J.R., Naggert J.K.,  
 RA Nishina P.M.;  
 RT "Mf1p, a gene encoding a frizzled related protein, is mutated in the  
 RT mouse retinal degeneration 6.";  
 RL Hum. Mol. Genet. 11:1879-1886(2002).  
 DR EMBL; AF469650; AAM69217.1;  
 DR MGD; MGI:238598; C1qntf5.  
 DR InterPro; IPR001073; C1q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1q; 1.  
 DR Pfam; PF01391; Collagen; 2.  
 DR PRINTS; PR00007; COMPLEMENTC1Q.  
 DR SMART; SM00110; C1Q; 1.  
 SQ SEQUENCE 243 AA; 25436 MW; 9F4D5804349791D9 CRC64;

Query Match 12.3%; Score 202; DB 11; Length 243;  
 Best Local Similarity 32.3%; Pred. No. 2.1e-09;  
 Matches 63; Conservative 17; Mismatches 81; Indels 34; Gaps 6;

```
QY 122 GAPHYALGAPGATFSGLYLVADADADAPA--RGPBPAD-----PEPRSAF 163
DB 57 GAP-----GAPGEGEGE-----GRPGLPGRGEPFRGAGPMGAIAPAGECSSVPRSAF 106
QY 164 SAARFSTSLVGSADGPPRRHQPLAFDFEFVNIIGSDPFAAGVFRCLRGAYFFSFTLGKLP 223
DB 107 SAKRSSSRV-----PEPADTLPFPDRVVLNBOGHVPTTGKFTCOVGVYFPA-VHAIVY 160
QY 224 RKTLSVYKLMKNRDEVQAMTYDDGASRRRMOQSVMALRRGPAVWLISHDHGYGAYSN 283
DB 161 RASLQGDVLXKNGSISFFQYFGCWKPKAPASLSGGAHVRLPEPDQVAVGVGVYIGIVAS 220
QY 284 HGKTYTSGFLVYPD 298
DB 221 IKTDSTFSGFLVYSD 235

RESULT 13
QY 09D8U4 PRELIMINARY; PRT; 294 AA.
AC 09D8U4;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE 1810033K05RIK protein (RIKEN cDNA 1810033K05 gene).
GN 1810033K05RIK.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=pancreas;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi S., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamamaki I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Frieschmann W., Gaasterland T., Giesi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Mashio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Batsch G.,
RA Blake J., Boiteil D., Bojunga N., Carinici P., de Bonaldi M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guelincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashiki Y.
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DB EMBL; AK007683; BAB25187.1; -
DB EMBL; BC030324; AAH30324.1; -
DB MGD; MGI:1916433; 1810033K05RIK.
DR InterPro; IPR001073; Clq.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF00386; Clq; 1.
DR Pfam; PF01391; Collagen; 2.
DR PRINTS; PR00007; COMPLEMENTC1Q.
DR PROSITE; PS01113; Clq; 1.
SQ SEQUENCE 294 AA; 30865 MW; 6D3905AE7C19E6FA CRC64;

Query Match 12.2%; Score 200; DB 11; Length 294;
Best Local Similarity 34.9%; Pred. No. 4e-09;
Matches 53; Conservative 30; Mismatches 45; Indels 24; Gaps 8;
```

```
QY 5 PGP---GSSELSRASFSA--RTPLFEGTSEMAVTFDKYVYNIIGDFVATGQFRGVPA 59
DB 147 PGPSCSGSRKSAFSAFVAVTSYPRE---RLPIKFDKILMBGGHYNASGKAFVCSVEI 203
QY 60 YFFSFTAGKAPKHSLSVNLVNRDEVQALAFDEQRPGARRAASOSAMLOLDYGDVWLR 119
DB 204 YFTYDTITLA-NKHLAIGLVN-GGYRIRTFD--ANTGNHVASGSTLLALKEGDEVWLQ 259
QY 120 LHGA-----PHYALGAPGATFSGLYLVAD 143
DB 260 IFYSEQNGLFYDPYWT---DSLFTGFLIYAD 287

RESULT 14
QY 095JD7 PRELIMINARY; PRT; 243 AA.
AC 095JD7;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Adiponectin.
GN APM1.
OS Macaca mulatta (rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxId=9544;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=adipose tissue;
RX MEDLINE=21232234; PubMed=11334417;
RA Hotta K., Funahashi T., Bodkin N.L., Ortmeier H.K., Arita Y.,
RA Hansen B.C., Matsuzawa Y.;
RT "Circulating concentrations of the adipocyte protein adiponectin are
RT decreased in parallel with reduced insulin sensitivity during the
RT progression to type 2 diabetes in rhesus monkeys.";
RL Diabetes 50:1126-1133(2001).
DB EMBL; AF040407; AAK92202.1; -
DR InterPro; IPR001073; Clq.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF00386; Clq; 1.
DR Pfam; PF01391; Collagen; 1.
DR PRINTS; PR00007; COMPLEMENTC1Q.
DR PRODOM; PD000007; Collagen; 1.
DR SMART; SM00110; Clq; 1.
DR PROSITE; PS01113; Clq; 1.
SQ SEQUENCE 243 AA; 26264 MW; 49A45DAF2B4613FD CRC64;

Query Match 11.8%; Score 193; DB 6; Length 243;
Best Local Similarity 31.3%; Pred. No. 1.2e-08;
Matches 47; Conservative 30; Mismatches 55; Indels 18; Gaps 6;

QY 5 PGPSSSELSRASFSAARTPLEGTSEMAVTFDKYVYNIIGDFVATGQFRGVPAFFSF 64
DB 103 PEGGAVYTRSAFSAFVGLFTVTV--VPMPIRFTKIFYNQNMHYGSTGKHCINPGLYFAY 161
QY 65 TAGKAPH-----KSLSVMLVNRDEVQALAFDEQRPGARRAASOSAMLOLDYGDVWLR 119
DB 162 -----HITVVKQDVKVSLEFK-KDKAMLFTYDQYQENNVQ--ASGSVLLHLEVGQVWLQ 213
QY 120 LHGAH-----YALGAPGATFSGLYLVADAD 145
DB 214 VYGEGERNGLVADNDNDSTFTGFLLYHDTN 243

RESULT 15
QY 08BRW2 PRELIMINARY; PRT; 247 AA.
AC 08BRW2;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
```

DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)  
DE Adipocyte complement related protein of 30 kDa.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The PANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs";  
RL Nature 420:563-573(2002).  
DR EMBL; AK041214; BAC30866.1; -  
SQ SEQUENCE 247 AA; 26751 MW; 0D3FA64C789CAEF3 CRC64;

Query Match 11.5%; Score 189; DB 11; Length 247;  
Best Local Similarity 33.5%; Pred. No. 2.8e-08;  
Matches 53; Conservative 24; Mismatches 73; Indels 8; Gaps 5;  
Qy 146 ADAPAGPPAPP---EPRSAFSAATRSIVGSDAGPGRHOPLAFDTEFVNIGGDFDA 201  
Db 91 AEGP-RGFPGTGRKGGEALYYRSAFSVGLFTRVTPNVPPIRFKIFVQGNHYDGS 149  
Qy 202 AGVPRCLPGAVFSGFTLGKLPKRTLSVKLMKNRDEVOAMTYDDGASRRREMOSQVMLA 261  
Db 150 TGFYCNIPGLYFYSYHI-TVMKDVKYSLEK-KDKAVLFTYDQYQEKKNVDQASGSVLLH 207  
Qy 262 LRRGDAVWLISH-DHDGAGVSNHGKXITFSGELVYPD 298  
Db 208 LEVGDQVWLQYVGDDHNGLYADNVNDSTFTGFLLYHD 245

Search completed: January 12, 2004, 08:19:08  
Job time : 44.1078 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:09 ; Search time 47.1687 Seconds

(without alignments)  
1053.272 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_17\_329

Perfect score: 1639  
Sequence: 1 LGTRPGCGSSSLRSPAFSAAR.....LVYEDLAPAPPGIGASLL 313

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: A\_Geneseq\_19Jun03:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	1639	100.0	329	22 AAB61606 Human ZACRP4. Hom
2	1620	98.8	329	23 AAB61606 Human novel secret
3	1579	96.3	329	22 AAB61424 Monkey MANCO 245 p
4	1480	90.3	348	22 AAB61423 Human MANCO 245 pr
5	1476	90.1	334	22 AAB61466 Human MANCO 245 ma
6	1453	88.7	299	23 AAB61489 Murine MANCO 245 ex
7	1358.5	82.9	284	23 AAB61477 Human MANCO 245 C
8	809	49.4	199	23 AAB61488 Monkey MANCO 245 C
9	760	46.4	192	22 AAB61479 Human MANCO 245 cy

10	707.5	43.2	221	23 AAB61473 Adipocyte compleme
11	683	41.7	134	22 AAB61473 Monkey MANCO 245 C
12	664	40.5	127	22 AAB61478 Human secreted pro
13	646	39.4	126	22 AAB61469 Human MANCO 245 C1
14	639	39.0	126	22 AAB61489 Murine MANCO 245 C
15	635	38.7	125	22 AAB61477 Human MANCO 245 ex
16	622	37.9	126	22 AAB61472 Monkey MANCO 245 C
17	618	37.7	123	21 AAB61489 Human ORFX ORF1953
18	609	37.2	117	22 AAB61470 Human MANCO 245 C1
19	552	33.7	130	22 AAB61468 Mature monkey MANC
20	507	30.9	133	23 AAB61486 Adipocyte compleme
21	237.5	14.5	252	22 AAB61472 Human ZACRP5. Hom
22	237.5	14.5	252	23 AAB61472 Human molecule for
23	237.5	14.5	252	23 AAB61472 Human secreted pro
24	237.5	14.5	252	23 AAB61472 Human polypeptide
25	231	14.1	800	22 AAB61470 Human protein SEQ
26	230	14.0	202	22 AAB61470 Human polypeptide
27	230	14.0	709	23 AAB61470 Human pancreatic c
28	230	14.0	710	23 AAB61470 Human genesc metab
29	230	14.0	746	20 AAB61470 Human lung tumour
30	230	14.0	746	21 AAB61470 Human lung tumour
31	230	14.0	746	21 AAB61470 Human lung tumour
32	230	14.0	908	22 AAB61470 Novel human protei
33	230	14.0	908	22 AAB61470 Novel human protei
34	230	14.0	957	22 AAB61470 Novel human protei
35	230	14.0	958	22 AAB61470 Novel human protei
36	230	14.0	992	22 AAB61470 Novel human protei
37	230	14.0	993	22 AAB61470 Novel human protei
38	230	14.0	1042	22 AAB61470 Novel human protei
39	230	14.0	1043	22 AAB61470 Novel human protei
40	230	14.0	1091	22 AAB61470 Novel human protei
41	230	14.0	1092	22 AAB61470 Novel human protei
42	230	14.0	1126	22 AAB61470 Novel human protei
43	230	14.0	1127	22 AAB61470 Novel human protei
44	220.5	13.5	281	20 AAB61470 Mouse adipocyte-sp
45	220.5	13.5	281	21 AAB61470 Amino acid sequenc

# ALIGNMENTS

RESULT 1	AA61606 standard; Protein; 329 AA.
ID	AA61606
XX	AA61606
AC	AA61606
XX	05-APR-2001 (first entry)
DT	05-APR-2001 (first entry)
XX	Human ZACRP4.
DE	Human ZACRP4.
XX	Human; zacr4; complement factor C1q domain; chromosome 11q11;
KW	energy balance; cellular metabolic reaction; autocrine factor;
KM	development; cell proliferation; differentiation; cell survival.
XX	Human sapiens.
OS	Human sapiens.
XX	Key
FT	Peptide
FT	Domain
FT	Domain
FT	Domain
FT	Domain
XX	Location/Qualifiers
XX	1..16
XX	17..159
XX	/label= Signal_peptide
XX	/label= C1q_domain_#1
XX	160..328
XX	/label= C1q_domain_#2
XX	WO200102565-A2.
XX	11-JAN-2001.
XX	28-JUN-2000; 2000WO-US17692.
XX	01-JUL-1999; 99US-0346502.

PA (Zymo) ZYMOGENETICS INC.  
 XX  
 PI Holloway JL, Lok S;  
 XX  
 DR WPI; 2001-138140/14.  
 XX N-PSDB; AAF28672.  
 PT Novel secreted protein ZACRP4 polypeptides having tandem C1q globular  
 PT domains, useful for studying cell-cell communication and regulation of  
 PT cellular processes -  
 XX  
 PS Claim 1; Page 77-78; 82pp; English.  
 XX  
 CC The present sequence is human ZACRP4 protein. ZACRP4 protein has two  
 CC complement factor C1q domains. The ZACRP4 gene is located on human  
 CC chromosome 11q11. The ZACRP4 coding sequence and protein have a number of  
 CC uses described in the specification, including, modulation of energy  
 CC balance and cellular metabolic reactions in mammals. In addition, ZACRP4  
 CC protein is useful as an autocrine factor, particularly during  
 CC development, in mediating the processes of an organism, in regulating  
 CC cellular processes such as cell proliferation and/or differentiation,  
 CC cell survival and energy balance.  
 CC  
 XX  
 SQ Sequence 329 AA;  
 Query Match 100.0%; Score 1639; DB 22; Length 329;  
 Best Local Similarity 100.0%; Pred. No. 4e-152;  
 Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LGPTGPGSSSLRASFSAARTTPLEGTSEMAVTFPKVYVNIIGDDVDVATGFRCHVPGAY 60  
 DB 17 LGPTGPGSSSLRASFSAARTTPLEGTSEMAVTFPKVYVNIIGDDVDVATGFRCHVPGAY 76  
 QY 61 FFSFTAGAPKHSLSVMLVRNDEVQALAFDEORRPGARRASQSAMQLDYGDTVMRL 120  
 DB 77 FFSFTAGAPKHSLSVMLVRNDEVQALAFDEORRPGARRASQSAMQLDYGDTVMRL 136  
 QY 121 HGAPHYALGAPGATSGVLYVADADADAPARGPAPPEPRSAFSAARTSLVSGDAGGCP 180  
 DB 137 HGAPHYALGAPGATSGVLYVADADADAPARGPAPPEPRSAFSAARTSLVSGDAGGCP 196  
 QY 181 RHQPLAFDTEFVNIIGDDFPAAAGVRCRPLPGVFFSFTLGKLPKRTL SVKMKNDEVQ 240  
 DB 197 RHQPLAFDTEFVNIIGDDFPAAAGVRCRPLPGVFFSFTLGKLPKRTL SVKMKNDEVQ 256  
 QY 241 MYDDGASRRRMOSSQVWALRRGDPAVWLSDHDGCVAGYVNIIGDDVDVATGFRCHVPGAY 300  
 DB 257 MYDDGASRRRMOSSQVWALRRGDPAVWLSDHDGCVAGYVNIIGDDVDVATGFRCHVPGAY 316  
 QY 301 PAAPGLGASELL 313  
 DB 317 PAAPGLGASELL 329  
 RESULT 2  
 ID ABG79643  
 XX ABG79643 standard; Protein; 329 AA.  
 AC ABG79643;  
 XX  
 DT 15-NOV-2002 (first entry)  
 XX  
 DE Human novel secreted protein SECP19, Incyte ID No. 931619CD1.  
 XX  
 KW Human; SECP, secreted protein; cell proliferative disorder;  
 KW actinic keratosis; arteriosclerosis; bursitis; hepatitis; cancer;  
 KW autoimmune disorder; inflammatory disorder; AIDS; asthma; allergy;  
 KW acquired immunodeficiency syndrome; anaemia; atopic dermatitis;  
 KW cardiovascular disorder; congestive heart failure; vascular tumour;  
 KW ischaemic heart disease; myocardial infarction; epilepsy; stroke;  
 KW hypertensive heart disease; neurological disorder; cerebral neoplasm;  
 KW Alzheimer's disease; developmental disorder; renal tubular acidosis;  
 KW Cushing's syndrome; Duchenne muscular dystrophy; hypothyroidism;

KW Becker muscular dystrophy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200262841-A2.  
 XX  
 PD 15-AUG-2002.  
 XX  
 PF 28-JAN-2002; 2002WO-US02616.  
 XX  
 PR 02-FEB-2001; 2001US-266195P.  
 PR 08-FEB-2001; 2001US-267924P.  
 PR 09-FEB-2001; 2001US-267816P.  
 PR 09-FEB-2001; 2001US-268112P.  
 PR 26-FEB-2001; 2001US-271639P.  
 PR 07-SEP-2001; 2001US-317818P.  
 PR 21-DEC-2001; 2001US-343533P.  
 XX  
 FA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Tang TY, Yue H, Gandhi AR, Yao MG, Warren BA, Ding L, Duggan BM;  
 PI Xu Y, Yang J, Thangavelu K, Lai PG, Honchell CD, Walla NK, Lee S;  
 PI Lee EA, Richardson TW, Baughn MR, Elliott VS;  
 XX  
 DR WPI; 2002-657522/70.  
 XX N-PSDB; ABS64954.  
 PT New human secreted proteins and nucleic acids useful in diagnosing,  
 PT treating and preventing cell proliferative, autoimmune/inflammatory,  
 PT cardiovascular, neurological, and developmental disorders -  
 XX  
 PS Claim 1; Page 140; 158pp; English.  
 XX  
 CC The invention relates to twenty four human secreted proteins  
 CC (SECP1-24), proteins 90% identical to them and active fragments of them.  
 CC Also included are nucleic acids encoding the SECP proteins, a recombinant  
 CC polynucleotide comprising a promoter sequence operably linked to the  
 CC nucleic acid, a cell transformed with the recombinant polynucleotide, an  
 CC transgenic organism comprising the recombinant polynucleotide, an  
 CC anti-SECP antibody, and screening for ant/agonists and modulators of  
 CC SECP function or expression. The SECP proteins and nucleic acids are  
 CC useful in the diagnosis, treatment and prevention of cell proliferative  
 CC (e.g. actinic keratosis, arteriosclerosis, bursitis, hepatitis or  
 CC cancer), autoimmune/inflammatory (e.g. AIDS (acquired immunodeficiency  
 CC syndrome), asthma, allergies or atopic dermatitis),  
 CC cardiovascular (e.g. congestive heart failure, ischaemic heart disease,  
 CC myocardial infarction, hypertensive heart disease, or vascular tumours),  
 CC neurological (e.g. epilepsy, stroke, cerebral neoplasms, or Alzheimer's  
 CC disease), and developmental (e.g. renal tubular acidosis, Cushing's  
 CC syndrome, Duchenne and Becker muscular dystrophy, or hypothyroidism)  
 CC disorders. Many other diseases and disorders are listed in the  
 CC specification. These may also be used in assessing the effects of  
 CC exogenous compounds on the expression of nucleic acid and amino acid  
 CC sequences of the secreted proteins. The present sequence represents a  
 CC SECP protein of the invention.  
 CC  
 XX  
 SQ Sequence 329 AA;  
 Query Match 98.8%; Score 1620; DB 23; Length 329;  
 Best Local Similarity 99.4%; Pred. No. 2.9e-150;  
 Matches 311; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 LGPTGPGSSSLRASFSAARTTPLEGTSEMAVTFPKVYVNIIGDDVDVATGFRCHVPGAY 60  
 DB 17 LGPTGPGSSSLRASFSAARTTPLEGTSEMAVTFPKVYVNIIGDDVDVATGFRCHVPGAY 76  
 QY 61 FFSFTAGAPKHSLSVMLVRNDEVQALAFDEORRPGARRASQSAMQLDYGDTVMRL 120  
 DB 77 FFSFTAGAPKHSLSVMLVRNDEVQALAFDEORRPGARRASQSAMQLDYGDTVMRL 136  
 QY 121 HGAPHYALGAPGATSGVLYVADADADAPARGPAPPEPRSAFSAARTSLVSGDAGGCP 180  
 DB 137 LGAPHYALGAPGATSGVLYVADADADAPARGPAPPEPRSAFSAARTSLVSGDAGGCP 196



QY 181 RHQPLADTEFVNIGGDFDAAGVFRCLPGAYFFSTLGLPKRTLSVKLMKRDVQA 240  
DB 197 RHQPLADTEFVNIGGDFDAAGVFRCLPGAYFFSTLGLPKRTLSVKLMKRDVQA 256  
QY 241 MIYDDGASRRREMOSQSVMLALRRGDVAVMLSHDHGAGVSNHGKITYTFSGLVYPDLA 300  
DB 257 MIYDDGASRRREMOSQSVMLALRRGDVAVMLSHDHGAGVSNHGKITYTFSGLVYPDLA 316  
QY 301 PAAPPGIGASELL 313  
DB 317 PAAPPGIGASELL 329

RESULT 3  
AAB61424  
ID AAB61424 standard; protein; 329 AA.  
XX  
AC AAB61424;  
XX  
DT 04-APR-2001 (first entry)  
XX  
DE Monkey MANGO 245 protein.  
XX  
KM TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
KM autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
KM pancreatic; skeletal; muscle.  
XX  
OS Cataract sp..  
XX  
PN WO200100672-A1.  
XX  
PD 04-JAN-2001.  
XX  
PF 29-JUN-2000; 2000WO-US18184.  
XX  
PR 29-JUN-1999; 99US-0342687.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;  
XX  
DR WPI; 2001-050127/06.  
XX  
PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
PT disorders (e.g. jaundice) -  
XX  
PS Claim 1; Fig 25; 262pp; English.  
XX  
CC The present invention relates to cDNAs encoding TANGO 244,  
CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
CC The nucleic acids, proteins and protein modulators are useful for  
CC treating colonic disorders, inflammatory diseases, tumors,  
CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
CC allergic diseases, cardiovascular diseases, brain disorders,  
CC degenerative diseases placental, pancreatic, skeletal and muscle  
CC disorders.  
XX  
SQ Sequence 329 AA;  
Query Match 96.3%; Score 1579; DB 22; Length 329;  
Best Local Similarity 96.5%; Pred. No. 3.1e-146;  
Matches 302; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 1 LGPTPGGSSSELRSASARTTPLEGTSEMAVTFDKYVYVIGDFDVATGQFCRVPAGAY 60  
DB 17 LGPTPGGSSSELRSASARTTPLEGTSEMAVTFDKYVYVIGDFDVATGQFCRVPAGAY 76  
QY 61 FFSTTACKAPHKSLSVMLVNRNDEVQALAPDEORRPARPARASOSAMLQIDYGTWLR 120  
DB 77 FFSTTACKAPHKSLSVMLVNRNDEVQALAPDEORRPARPARASOSAMLQIDYGTWLR 136

QY 121 HGAPVYALGAPATFSSGYLYADADADAPARPPAPPEPRPSAFAAARTSLVGSAGPGP 180  
DB 137 HGAPVYALGAPATFSSGYLYADADADAPARPPAPPEPRPSAFAAARTSLVGSAGPGP 196  
QY 181 RHQPLADTEFVNIGGDFDAAGVFRCLPGAYFFSTLGLPKRTLSVKLMKRDVQA 240  
DB 197 RHQPLADTEFVNIGGDFDAAGVFRCLPGAYFFSTLGLPKRTLSVKLMKRDVQA 256  
QY 241 MIYDDGASRRREMOSQSVMLALRRGDVAVMLSHDHGAGVSNHGKITYTFSGLVYPDLA 300  
DB 257 MIYDDGASRRREMOSQSVMLALRRGDVAVMLSHDHGAGVSNHGKITYTFSGLVYPDLA 316  
QY 301 PAAPPGIGASELL 313  
DB 317 PAAPPGIGASELL 329

RESULT 4  
AAB61423  
ID AAB61423 standard; protein; 348 AA.  
XX  
AC AAB61423;  
XX  
DT 04-APR-2001 (first entry)  
XX  
DE Human MANGO 245 protein.  
XX  
KM TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
KM autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
KM pancreatic; skeletal; muscle.  
XX  
OS Homo sapiens.  
XX  
PN WO200100672-A1.  
XX  
PD 04-JAN-2001.  
XX  
PF 29-JUN-2000; 2000WO-US18184.  
XX  
PR 29-JUN-1999; 99US-0342687.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;  
XX  
DR WPI; 2001-050127/06.  
XX  
PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
PT disorders (e.g. jaundice) -  
XX  
PS Claim 1; Fig 23; 262pp; English.  
XX  
CC The present invention relates to cDNAs encoding TANGO 244,  
CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
CC The nucleic acids, proteins and protein modulators are useful for  
CC treating colonic disorders, inflammatory diseases, tumors,  
CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
CC allergic diseases, cardiovascular diseases, brain disorders,  
CC degenerative diseases placental, pancreatic, skeletal and muscle  
CC disorders.  
XX  
SQ Sequence 348 AA;  
Query Match 90.3%; Score 1480; DB 22; Length 348;  
Best Local Similarity 99.6%; Pred. No. 1.7e-136;  
Matches 283; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LGPTPGGSSSELRSASARTTPLEGTSEMAVTFDKYVYVIGDFDVATGQFCRVPAGAY 60  
DB 17 LGPTPGGSSSELRSASARTTPLEGTSEMAVTFDKYVYVIGDFDVATGQFCRVPAGAY 76

QY 61 FFSEFTAGKAPHKSLVSVLVRNDEVOALAFDEQRPGARRAASQSAAMLQIDYGDYTWLRL 120  
DB 77 FFSEFTAGKAPHKSLVSVLVRNDEVOALAFDEQRPGARRAASQSAAMLQIDYGDYTWLRL 136  
QY 121 HGAPHYALGAPGATFSGYLVYADADADAPARGPAPPEPRSAFSAARTSLVGSADAGP 180  
DB 137 HGAPQYALGAPGATFSGYLVYADADADAPARGPAPPEPRSAFSAARTSLVGSADAGP 196  
QY 181 RHOPLAPDTEFEVNIIGDFDAAGVFRCLPGAYFFSFTLGLPKRTLSVLMKRNDEVOA 240  
DB 197 RHOPLAPDTEFEVNIIGDFDAAGVFRCLPGAYFFSFTLGLPKRTLSVLMKRNDEVOA 256  
QY 241 MIYDDGASRRREMOSQSVMLALRRGDAVWLLSHDHGCGAYSNH 284  
DB 257 MIYDDGASRRREMOSQSVMLALRRGDAVWLLSHDHGCGAYSNH 300

RESULT 5  
AAB61466  
ID AAB61466 standard; protein; 334 AA.  
XX  
AC AAB61466;  
XX  
DT 04-APR-2001 (first entry)  
XX  
DE Human MANGO 245 mature protein.  
XX  
KM TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
KM autoimmunity; allergy; cardiovascular; brain; degenerative; placental;  
KM pancreatic; skeletal; muscle.  
XX  
OS Homo sapiens.  
XX  
EN WO200100672-A1.  
XX  
PD 04-JAN-2001.  
XX  
PF 29-JUN-2000; 2000WO-US18184.  
XX  
PR 29-JUN-1999; 99US-0342687.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Holtzman DA, Barnes TM, Frazer CC, Sharp JD;  
XX  
DR WPI, 2001-050127/06.  
XX  
PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
PT disorders (e.g. jaundice) -  
XX  
PS Disclosure; Page 242; 262pp; English.  
XX  
CC The present invention relates to cDNAs encoding TANGO 244,  
CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
CC The nucleic acids, proteins and protein modulators are useful for  
CC treating colonic disorders, inflammatory diseases, tumors,  
CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
CC allergic diseases, cardiovascular diseases, brain disorders,  
CC degenerative diseases placental, pancreatic, skeletal and muscle  
CC disorders.  
XX  
SQ Sequence 334 AA;  
Query Match 90.1%; Score 1476; DB 22; Length 334;  
Best Local Similarity 99.3%; Pred. No. 3.9e-136;  
Matches 282; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 LGPTGPGSSSELRSAPARTTPLBGTSEMAVTFDKVYVNIIGDFDVAATGRCRVPAY 60  
DB 3 LGPTGPGSSSELRSAPARTTPLBGTSEMAVTFDKVYVNIIGDFDVAATGRCRVPAY 62

QY 61 FFSEFTAGKAPHKSLVSVLVRNDEVOALAFDEQRPGARRAASQSAAMLQIDYGDYTWLRL 120  
DB 63 FFSEFTAGKAPHKSLVSVLVRNDEVOALAFDEQRPGARRAASQSAAMLQIDYGDYTWLRL 122  
QY 121 HGAPHYALGAPGATFSGYLVYADADADAPARGPAPPEPRSAFSAARTSLVGSADAGP 180  
DB 123 HGAPQYALGAPGATFSGYLVYADADADAPARGPAPPEPRSAFSAARTSLVGSADAGP 182  
QY 181 RHOPLAPDTEFEVNIIGDFDAAGVFRCLPGAYFFSFTLGLPKRTLSVLMKRNDEVOA 240  
DB 183 RHOPLAPDTEFEVNIIGDFDAAGVFRCLPGAYFFSFTLGLPKRTLSVLMKRNDEVOA 242  
QY 241 MIYDDGASRRREMOSQSVMLALRRGDAVWLLSHDHGCGAYSNH 284  
DB 243 MIYDDGASRRREMOSQSVMLALRRGDAVWLLSHDHGCGAYSNH 286

RESULT 6  
ABG70385  
ID ABG70385 standard; Protein; 299 AA.  
XX  
AC ABG70385;  
XX  
DT 05-NOV-2002 (first entry)  
XX  
DE Adipocyte complement-related C1q Tumour Necrosis Factor-like protein #3.  
XX  
KM Human; NOVX; NOVX-associated disorder; cardiomyopathy; atherosclerosis;  
KM cell signal processing; metabolic pathway modulation; metabolic disorder;  
KM obesity; diabetes; infectious disease; neurodegenerative disorder; acne;  
KM Alzheimer's disease; Parkinson's disease; immune disorder; cancer;  
KM haematopoietic disorder; cirrhosis; pancreatitis; learning defect;  
KM memory defect; infertility; congenital heart defect; hair growth;  
KM pigmentation disorder; endocrine disorder; respiratory disease; health;  
KM gastro-intestinal disease; reproductive; neurological disease;  
KM bone marrow transplantation; endocrine disease; allergy; inflammation;  
KM nephrological disorder; urinary system disorder; age-related disorder;  
KM neuropsychiatric disorder; EGF-related protein; SCUB1; TEN-M4;  
KM adipocyte complement-related C1q tumour necrosis factor; cut at first;  
KM beta adrenergic receptor kinase; EphA6/enk-2; glucose transporter;  
KM type Ia membrane suah1-containing domain; butyrophilin;  
KM type Ia membrane-sushi domain containing.  
XX  
OS Homo sapiens.  
XX  
EN WO200257453-A2.  
XX  
PD 25-JUL-2002.  
XX  
PF 19-DEC-2001; 2001WO-US50331.  
XX  
PR 19-DEC-2000; 2000US-265704P.  
PR 20-DEC-2000; 2000US-257314P.  
PR 02-MAY-2001; 2001US-286153P.  
PR 29-MAY-2001; 2001US-294075P.  
PR 24-JUL-2001; 2001US-307506P.  
PR 10-AUG-2001; 2001US-311590P.  
PR 10-AUG-2001; 2001US-311613P.  
PR 29-AUG-2001; 2001US-315617P.  
PR 14-SEP-2001; 2001US-322358P.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Gangoli EA, Paturajan M, Vernet CAM, Malpankar UM, Kekuda R;  
PI Stone DJ, Anderson D, Shimkets RA, Burgess CE, Zernhusen BD, Liu X;  
PI Spytek KA, Casman SJ, Boidog FL, Smithson G, Li L, Ji W;  
XX  
DR WPI, 2002-590744/63.  
XX  
N-PSDB; ABS52097.  
XX  
PT Novel isolated NOVX polypeptide useful for treating cardiomyopathy,  
PT atherosclerosis, metabolic disorders, diabetes, obesity, infectious

Db 77 FFSTAGKAPHKSLVLRNDEVQALAFDEQRRPGARRAASQSAMLQLDYGDTVWLRL 13



QY 261 ALRGDAVWLISHDHGCGAYSNH 284  
 DB 121 ALRGDAVWLISHDHGCGAYSNH 144

RESULT 10  
 ABG70383  
 ID ABG70383 standard; Protein: 221 AA.  
 AC AEG70383;  
 DT 05-NOV-2002 (first entry)  
 DE Adipocyte complement-related C1q Tumour Necrosis Factor-like protein #1.  
 XX Human; NOX; NOX-associated disorder; cardiomyopathy; atherosclerosis;  
 XX cell signal processing; metabolic pathway modulation; metabolic disorder;  
 XX obesity; diabetes; infectious disease; neurodegenerative disorder; acne;  
 XX Alzheimer's disease; Parkinson's disease; immune disorder; cancer;  
 XX haematopoietic disorder; cirrhosis; pancreatitis; learning defect;  
 XX memory defect; infertility; congenital heart defect; hair growth;  
 XX pigmentation disorder; endocrine disorder; respiratory disease; health;  
 XX gastro-intestinal disease; reproductive; neurological disease;  
 XX bone marrow transplantation; endocrine disease; allergy; inflammation;  
 XX nephrological disorder; urinary system disorder; age-related disorder;  
 XX neuropsychiatric disorder; BGF-related protein; SCUB1; TEN-M4;  
 XX adipocyte complement-related C1q tumour necrosis factor; out at first;  
 XX beta adrenergic receptor kinase; EphA6/ehk-2; glucose transporter;  
 XX type Ia membrane sushi-containing domain; butyrophilin;  
 XX type Ia membrane sushi domain containing.  
 OS Homo sapiens.  
 PN WO200257453-A2.  
 PD 25-JUL-2002.  
 XX 19-DEC-2001; 2001MO-US50331.  
 PF 19-DEC-2000; 2000US-265704P.  
 PR 20-DEC-2000; 2000US-257314P.  
 PR 02-MAY-2001; 2001US-288153P.  
 PR 29-MAY-2001; 2001US-294075P.  
 PR 24-JUL-2001; 2001US-307506P.  
 PR 10-AUG-2001; 2001US-311590P.  
 PR 10-AUG-2001; 2001US-311613P.  
 PR 29-AUG-2001; 2001US-315617P.  
 PR 14-SEP-2001; 2001US-322358P.  
 PA (CURA-) CURAGEN CORP.  
 XX Gangoli EA, Paturajan M, Vernet CAM, Malyankar UM, Kekuda R;  
 PI Stone DJ, Anderson D, Shimkets RA, Burgess CE, Zernhusen BD, Liu X;  
 PI Spytek KA, Caeman SJ, Boldog FL, Smithson G, Li L, Ji W;  
 XX WPI; 2002-590744/63.  
 DR N-PSDB; ABB52095.  
 DR Novel isolated NOX polypeptide useful for treating cardiomyopathy,  
 PT atherosclerosis, metabolic disorders, diabetes, obesity, infectious  
 PT disease, anorexia, neurodegenerative disorders, Alzheimer's disease or  
 PT cancer  
 XX Claim 1; Page 28; 318pp; English.

CC The present invention relates to new NOX polypeptides. The invention is  
 CC useful for treating or preventing a NOX-associated disorder such as  
 CC cardiomyopathy or atherosclerosis, where the disorder is related to cell  
 CC signal processing and metabolic pathway modulation in a subject,  
 CC preferably human. The invention is also useful for treating metabolic  
 CC disorders (e.g. obesity), diabetes, infectious disease, neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease), immune  
 CC disorders, haematopoietic disorders and various cancers. The molecules of

CC the invention are also useful for treating or preventing cirrhosis,  
 CC pancreatitis, learning and memory defects, infertility, congenital heart  
 CC defects, acne, hair growth, pigmentation disorders, endocrine disorders,  
 CC respiratory disease, gastro-intestinal diseases, reproductive, health,  
 CC neurological diseases, bone marrow transplantation, endocrine diseases,  
 CC allergy and inflammation, nephrological disorders, urinary system  
 CC disorders, neuropsychiatric disorders and age-related disorders.  
 CC The present amino acid sequence represents a NOX protein of the  
 CC invention.  
 CC  
 SQ Sequence 221 AA;  
 Query Match 43.2%; Score 707.5; DB 23; Length 221;  
 Best Local Similarity 49.5%; Pred. No. 4,8e-61;  
 Matches 153; Conservative 1; Mismatches 2; Indels 153; Gaps 1;

QY 5 PGRGSSELRSAPFAATPTLEGTSEMAVTPDKYVNIIGDFDPAATGQFRCRVGAFFSF 64  
 DB 66 PDGSSAMMSAFSAANTTPLEGTSEMAVTFDKY----- 99  
 QY 65 TAGKAPHKSLSVLVRNDEVQALAFDEQRPRGARRAASQAMLQDYGDTWLRHGAP 124  
 DB 100 ----- 99  
 QY 125 HVALGAPGATFSGYLVYADADADAPARCPPAPPEPRSAFSAARTSLVGSADGPPRRHP 184  
 DB 100 ----- 99  
 QY 185 LAFDTFVNIIGDFDPAAGVPRCLPGAYFSTTGLPRKTLVTKMKNRDEVQAMTYD 244  
 DB 100 -----VNIIGDFDPAAGVPRCLPGAYFSTTGLPRKTLVTKMKNRDEVQAMTYD 152  
 QY 245 DGASRRREMOSQSVMLALRGDAVWLISHDHGCGAYSNHGKXITFSGFLVYPDLAPAP 304  
 DB 153 DGASRRREMOSQSVMLALRGDAVWLISHDHGCGAYSNHGKXITFSGFLVYPDLAPAP 212  
 QY 305 PGLGASELL 313  
 DB 213 PGLGASELL 221

RESULT 11  
 AAB61473  
 ID AAB61473 standard; protein: 134 AA.  
 AC AAB61473;  
 XX 04-APR-2001 (first entry)  
 DT 04-APR-2001 (first entry)  
 XX Monkey MANGO 245 C1q domain #2.  
 DE TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
 XX autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
 KW pancreatic; skeletal; muscle.  
 OS Catarrhini sp..  
 XX WO200100672-A1.  
 PN 04-JUN-2001.  
 XX 29-JUN-2000; 2000WO-US18184.  
 PP 29-JUN-1999; 99US-0342687.  
 PR (MILL-) MILLENNIUM PHARM INC.  
 XX Holtzman DA, Barnes TM, Fraser CC, Sharp JD;  
 PI WPI; 2001-050127/06.  
 DR Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
 PT MANGO 245 proteins, useful in the treatment of inflammatory diseases

(e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver disorders (e.g. jaundice) -

Disclosure; Fig 28; 262pp; English.

CC The present invention relates to cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.

CC The nucleic acids, proteins and protein modulators are useful for treating colonic disorders, inflammatory diseases, tumors, renal disorders, liver disorders, lung disorders, autoimmune diseases, allergic diseases, cardiovascular diseases, brain disorders, degenerative diseases placental, pancreatic, skeletal and muscle disorders.

CC

XX Sequence 134 AA;

SQ

Query Match 41.7%; Score 683; DB 22; Length 134;  
Best Local Similarity 97.8%; Pred. No. 6,3e-59;  
Matches 131; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 162 AFSAAATRLVGSAGGPRPHQPLAFDTFVNIGDPPDAAGVRCRLPGAYFFSFTLTK 221  
1 AFSAAATRLVGSAGGPRPHQPLAFDTFVNIGDPPDAAGVRCRLPGAYFFSFTLTK 60

DB 222 LPKRTLSVKLMKNRDEVQAMLYDDGASRRRMOSSVWALRRGDVWLLSHDHGCGAY 281  
61 LPKRTLSVKLMKNRDEVQAMLYDDGASRRRMOSSVWALRRGDVWLLSHDHGCGAY 120

QY 282 SNHGKYYTFSGFLV 295  
121 SNHGKYYTFSGFLV 134

DB

RESULT 12  
ABBI1578  
ID ABBI1578 standard; peptide; 127 AA.

XX

AC ABBI1578;

XX

DT 11-JAN-2002 (first entry)

XX

DE Human secreted protein homologue, SEQ ID NO:1948.

XX

Human, cytokine; cell proliferation; cell differentiation; growth factor; haematopoiesis regulation; tissue growth; immunomodulator; activin; inhibin; chemotaxis; chemokinesis; thrombolytic; oncogenesis; proliferation; metastasis; cancer; tumour; haematopoietic disorder; myeloid cell disorder; lymphoid cell disorder; asthma; arthritis; chronic inflammatory condition; proliferative retinopathy; atherosclerosis; coronary heart disease; arterial ischaemia; bone disorder; osteoporosis; vascular growth disorder; tissue regeneration; wound healing; infection; immune disorder; cell culture; drug screening; gene therapy; antiinflammatory; antiaesthetic; antiarthritic; haemostatic; antiarteriosclerotic; cytoskeletal; osteoplastic; vasotropic; cardiant; virucide; antibacterial; antifungal; vulnetary; antilicer.

XX

OS Homo sapiens.

XX

PN WO200157188-A2.

PD 09-AUG-2001.

XX

PF 05-FEB-2001; 2001WO-US03800.

XX

PR 03-FEB-2000; 2000US-0496914.

XX

PR 27-APR-2000; 2000US-0560875.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Drmanac RT;

XX

DR WPI, 2001-457740/49.

DR N-PSDB; ABA08822.

XX

XX Human proteins and DNA encoding sequences useful for preventing, treating or ameliorating a medical condition in a mammalian subject

PT e.g. arthritis and cancer -

PT

PS Claim 20; Page 214-215; 1963pp; English.

XX

XX Sequences ABB10981-ABBI12330 represent 1350 novel human polypeptides, and sequences ABA08225-ABA09574 represent nucleic acids encoding them. The invention also relates to vectors and recombinant host cells comprising a nucleotide of the invention, methods of producing the novel polypeptides, antibodies against the polypeptides, methods of detecting the nucleotides or polypeptides in a sample, and methods of identifying compounds which bind to polypeptides of the invention. Although novel, many of the polypeptides of the invention have homology to known proteins, thereby giving an insight into their probable biological activities, and hence potential therapeutic applications. The polypeptides of the invention may have various activities, including cytokine, cell proliferation or cell differentiation activities; stem cell growth factor activity; haematopoiesis regulatory activity; tissue growth activity; immunomodulatory activity; activity- or inhibin-related activities; chemotactic or chemokinetic activities; haemostatic, thrombotic or thrombolytic activities; receptor or ligand activities; or may be involved in oncogenesis, cancer cell proliferation or metastasis. CC depending on their biological activities, polypeptides and nucleotides of the invention are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Such conditions include cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell disorders), chronic inflammatory conditions (e.g., asthma or arthritis), proliferative retinopathy, atherosclerosis, coronary heart disease, arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g., of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. CC Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, CC autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a novel human polypeptide of the invention.

CC

XX

SQ Sequence 127 AA;

Query Match 40.5%; Score 664; DB 22; Length 127;  
Best Local Similarity 99.2%; Pred. No. 4,3e-57;  
Matches 126; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 187 FDTFVNIGDPPDAAGVRCRLPGAYFFSFTLTKLPKRTLSVKLMKNRDEVQAMLYDDG 246  
1 FDTFVNIGDPPDAAGVRCRLPGAYFFSFTLTKLPKRTLSVKLMKNRDEVQAMLYDDG 60

DB 247 ASRRRMOSSVWALRRGDVWLLSHDHGCGAYSNHGKYYTFSGFLVYDPLAPAAPG 306  
61 SSRRRMOSSVWALRRGDVWLLSHDHGCGAYSNHGKYYTFSGFLVYDPLAPAAPG 120

QY 307 LGASELL 313  
121 LGASELL 127

DB

RESULT 13  
AAB61469  
ID AAB61469 standard; protein; 126 AA.

XX

AC AAB61469;

XX

DT 04-APR-2001 (first entry)

XX

DE	Human MANGO 245 Clq domain #1.
XX	
KM	TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;
KW	autoimmune; allergy; cardiovascular; brain; degenerative; placental;
KM	pancreatic; skeletal; muscle.
XX	
OS	Homo sapiens.
XX	
PN	WO200100672-A1.
XX	
PD	04-JAN-2001.
PF	29-JUN-2000; 2000MO-US18184.
XX	
PR	29-JUN-1999; 99US-0342687.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
XX	
P1	Holtzman DA, Barnes TM, Fraser CC, Sharp JD;
XX	
DR	WPI; 2001-050127/06.
XX	
PT	Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and
PT	MANGO 245 proteins; useful in the treatment of inflammatory diseases
PT	(e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver
PT	disorders (e.g. jaundice) -
XX	
PS	Disclosure; Fig 26; 262pp; English.
XX	
CC	The present invention relates to cDNAs encoding TANGO 244,
CC	TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.
CC	The nucleic acids, proteins and protein modulators are useful for
CC	treating colonic disorders, inflammatory diseases, tumors,
CC	renal disorders, liver disorders, lung disorders, autoimmune diseases,
CC	allergic diseases, cardiovascular diseases, brain disorders,
CC	degenerative diseases placental, pancreatic, skeletal and muscle
CC	disorders.
XX	
SQ	Sequence 126 AA:
Query Match	39.4%; Score 646; DB 22; Length 126;
Best Local Similarity	99.2%; Pred. No.2.5e-55;
Matches 125; Conservative	0; Mismatches 1; Indels 0; Gaps 0.
QY	15 AFSARTTPLEGTSMAVTFDKVYVNIGDFEDVATGQRRCRVPGAYFFSFTGAKPHKSL 74
DB	1 AFSAARTTPLEGTSMAVTFDKVYVNIGDPEVAIGQRCRVPGAYFFSFTGAKPHKSL 60
DY	75 SVMLVRNDEVQAALAFDEQRPGARRASQSAMTLGDYGDTVMRLHGAAPHYALGAPGAT 134
DB	61 SVMIVRNDEVQAALAFDEQRPGARRASQSAMTLGDYGDTVMRLHGAAPHYALGAPGAT 120
QY	135 FSGYLTV 140
DB	121 FSGYLTV 126
RESULT 14	
AAB61489	
KM	AAB61489 standard; proteain; 126 AA.
AC	AAB61489;
XX	
DT	04-APR-2001 (first entry)
XX	
DE	Murine MANGO 245 Clq domain.
XX	
KW	TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;
KM	autoimmune; allergy; cardiovascular; brain; degenerative; placental;
KW	pancreatic; skeletal; muscle.
XX	
OS	Mus musculus.
XX	

Query	Sequence	Matches	Score	DB	Length	Indels	Gaps	Score
QY 15	AFSAARTTPELGISEMAVTEDEKVTYVNIIGDQFDVATGQFRCHVPGAYFFSFTAGKAPKSL 74	39.0%;	639;	DB 22;	126;	0;	0;	74
DB 1	AFSAARTTPELGISEMAVTEDEKVTYVNIIGDQFDVATGQFRCHVPGAYFFSFTAGKAPKSL 60	39.0%;	639;	DB 22;	126;	0;	0;	60
QY 75	SVMLVNRNDEVOALAFEDORRPGARRAASOSAMQLDYGDVTWMLRIGAPHYATGAPGAT 134	39.0%;	639;	DB 22;	126;	0;	0;	134
DB 61	SVMLVNRNDEVOALAFEDORRPGARRAASOSAMQLDYGDVTWMLRIGAPHYATGAPGAT 120	39.0%;	639;	DB 22;	126;	0;	0;	120
QY 135	FSGYLV 140	39.0%;	639;	DB 22;	126;	0;	0;	140
DB 121	FSGYLV 126	39.0%;	639;	DB 22;	126;	0;	0;	126
QY 15	AFSAARTTPELGISEMAVTEDEKVTYVNIIGDQFDVATGQFRCHVPGAYFFSFTAGKAPKSL 74	39.0%;	639;	DB 22;	126;	0;	0;	74
DB 1	AFSAARTTPELGISEMAVTEDEKVTYVNIIGDQFDVATGQFRCHVPGAYFFSFTAGKAPKSL 60	39.0%;	639;	DB 22;	126;	0;	0;	60
QY 75	SVMLVNRNDEVOALAFEDORRPGARRAASOSAMQLDYGDVTWMLRIGAPHYATGAPGAT 134	39.0%;	639;	DB 22;	126;	0;	0;	134
DB 61	SVMLVNRNDEVOALAFEDORRPGARRAASOSAMQLDYGDVTWMLRIGAPHYATGAPGAT 120	39.0%;	639;	DB 22;	126;	0;	0;	120
QY 135	FSGYLV 140	39.0%;	639;	DB 22;	126;	0;	0;	140
DB 121	FSGYLV 126	39.0%;	639;	DB 22;	126;	0;	0;	126
QY 15	AFSAARTTPELGISEMAVTEDEKVTYVNIIGDQFDVATGQFRCHVPGAYFFSFTAGKAPKSL 74	39.0%;	639;	DB 22;	126;	0;	0;	74
DB 1	AFSAARTTPELGISEMAVTEDEKVTYVNIIGDQFDVATGQFRCHVPGAYFFSFTAGKAPKSL 60	39.0%;	639;	DB 22;	126;	0;	0;	60
QY 75	SVMLVNRNDEVOALAFEDORRPGARRAASOSAMQLDYGDVTWMLRIGAPHYATGAPGAT 134	39.0%;	639;	DB 22;	126;	0;	0;	134
DB 61	SVMLVNRNDEVOALAFEDORRPGARRAASOSAMQLDYGDVTWMLRIGAPHYATGAPGAT 120	39.0%;	639;	DB 22;	126;	0;	0;	120
QY 135	FSGYLV 140	39.0%;	639;	DB 22;	126;	0;	0;	140
DB 121	FSGYLV 126	39.0%;	639;	DB 22;	126;	0;	0;	126
QY 15	AFSAARTTPELGISEMAVTEDEKVTYVNIIGDQFDVATGQFRCHVPGAYFFSFTAGKAPKSL 74	39.0%;	639;	DB 22;	126;	0;	0;	74
DB 1	AFSAARTTPELGISEMAVTEDEKVTYVNIIGDQFDVATGQFRCHVPGAYFFSFTAGKAPKSL 60	39.0%;	639;	DB 22;	126;	0;	0;	60
QY 75	SVMLVNRNDEVOALAFEDORRPGARRAASOSAMQLDYGDVTWMLRIGAPHYATGAPGAT 134	39.0%;	639;	DB 22;	126;	0;	0;	134
DB 61	SVMLVNRNDEVOALAFEDORRPGARRAASOSAMQLDYGDVTWMLRIGAPHYATGAPGAT 120	39.0%;	639;	DB 22;	126;	0;	0;	120
QY 135	FSGYLV 140	39.0%;	639;	DB 22;	126;	0;	0;	140
DB 121	FSGYLV 126	39.0%;	639;	DB 22;	126;	0;	0;	126
QY 15	AFSAARTTPELGISEMAVTEDEKVTYVNIIGDQFDVATGQFRCHVPGAYFFSFTAGKAPKSL 74	39.0%;	639;	DB 22;	126;	0;	0;	74
DB 1	AFSAARTTPELGISEMAVTEDEKVTYVNIIGDQFDVATGQFRCHVPGAYFFSFTAGKAPKSL 60	39.0%;	639;	DB 22;	126;	0;	0;	60
QY 75	SVMLVNRNDEVOALAFEDORRPGARRAASOSAMQLDYGDVTWMLRIGAPHYATGAPGAT 134	39.0%;	639;	DB 22;	126;	0;	0;	134
DB 61	SVMLVNRNDEVOALAFEDORRPGARRAASOSAMQLDYGDVTWMLRIGAPHYATGAPGAT 120	39.0%;	639;	DB 22;	126;	0;	0;	120
QY 135	FSGYLV 140	39.0%;	639;	DB 22;	126;	0;	0;	140
DB 121	FSGYLV 126	39.0%;	639;	DB 22;	126;	0;	0;	126
QY 15	AFSAARTTPELGISEMAVTEDEKVTYVNIIGDQFDVATGQFRCHVPGAYFFSFTAGKAPKSL 74	39.0%;	639;	DB 22;	126;	0;	0;	74
DB 1	AFSAARTTPELGISEMAVTEDEKVTYVNIIGDQFDVATGQFRCHVPGAYFFSFTAGKAPKSL 60	39.0%;	639;	DB 22;	126;	0;	0;	60
QY 75	SVMLVNRNDEVOALAFEDORRPGARRAASOSAMQLDYGDVTWMLRIGAPHYATGAPGAT 134	39.0%;	639;	DB 22;	126;	0;	0;	134
DB 61	SVMLVNRNDEVOALAFEDORRPGARRAASOSAMQLDYGDVTWMLRIGAPHYATGAPGAT 120	39.0%;	639;	DB 22;	126;	0;	0;	120
QY 135	FSGYLV 140	39.0%;	639;	DB 22;	126;	0;	0;	140
DB 121	FSGYLV 12							

PA (MILL-) MILLENNIUM PHARM INC.

XX  
PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;

XX  
DR WPI; 2001-050127/06.

XX  
PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
PT disorders (e.g. jaundice) -

XX  
PS Disclosure; Page 246; 262pp; English.

CC  
CC The present invention relates to cDNAs encoding TANGO 244,

CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.

CC The nucleic acids, proteins and protein modulators are useful for  
CC treating colonic disorders, inflammatory diseases, tumors,  
CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
CC allergic diseases, cardiovascular diseases, brain disorders,  
CC degenerative diseases placental, pancreatic, skeletal and muscle  
CC disorders.

XX  
SQ Sequence 125 AA;

Query Match 38.7%; Score 635; DB 22; Length 125;

Best Local Similarity 100.0%; Pred. No. 2.9e-54;

Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LGPTPGSSSELRSAFSAARTTPLEGTSEMAVTFPKYVYVNIIGDFDVATGQFRCRVPGAY 60

Db 3 LGPTPGSSSELRSAFSAARTTPLEGTSEMAVTFPKYVYVNIIGDFDVATGQFRCRVPGAY 62

OY 61 FFSTFAGKAPKSLVNLVNRNDEVQALAFDEQRPGARRAASQSAAMLQLDYGDVTWLR 120

Db 63 FFSTFAGKAPKSLVNLVNRNDEVQALAFDEQRPGARRAASQSAAMLQLDYGDVTWLR 122

OY 121 HGA 123

Db 123 HGA 125

Search completed: January 12, 2004, 08:15:35  
Job time : 48.1687 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:15:44 ; Search time 38.1098 Seconds  
(without alignments)  
1655.023 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_17\_329

Perfect score: 1639  
Sequence: 1 LGPTPGSSSLRKAFAAAR.....LVYPLAPAPGAGL 313

Scoring table:  
Gapop 10.0 , Gapext 0.5

Searched: 747907 seqs, 201509753 residues

Total number of hits satisfying chosen parameters: 747907

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/ECT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/ECTS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1639	100.0	329	12	US-10-085-167-2
2	1633	99.6	329	12	US-10-236-055A-14
3	1514	92.4	326	12	US-10-236-055A-16
4	237.5	14.5	252	10	US-09-893-737-52
5	230	14.0	202	15	US-10-091-458-43
6	230	14.0	420	12	US-10-236-055A-2
7	230	14.0	746	10	US-09-738-973-185
8	230	14.0	746	10	US-09-854-133-185
9	230	14.0	746	15	US-10-144-649A-185
10	230	14.0	908	12	US-10-309-422-10
11	230	14.0	909	12	US-10-309-422-22
12	230	14.0	957	12	US-10-309-422-14
13	230	14.0	958	12	US-10-309-422-26
14	230	14.0	992	12	US-10-309-422-18
15	230	14.0	993	12	US-10-309-422-30

16	230	14.0	1042	12	US-10-309-422-8	Sequence 8, Appl
17	230	14.0	1043	12	US-10-309-422-20	Sequence 20, Appl
18	230	14.0	1091	12	US-10-309-422-12	Sequence 12, Appl
19	230	14.0	1092	12	US-10-309-422-24	Sequence 24, Appl
20	230	14.0	1126	12	US-10-309-422-16	Sequence 16, Appl
21	230	14.0	1127	12	US-10-309-422-28	Sequence 28, Appl
22	226	13.8	409	12	US-10-236-055A-4	Sequence 4, Appl
23	220.5	13.5	281	10	US-09-911-176B-44	Sequence 44, Appl
24	220.5	13.5	281	12	US-10-360-186-44	Sequence 44, Appl
25	220.5	13.5	281	12	US-10-392-531-44	Sequence 44, Appl
26	220.5	13.5	281	12	US-10-392-706-44	Sequence 44, Appl
27	220.5	13.5	281	15	US-10-385-015-5	Sequence 5, Appl
28	220.5	13.5	281	15	US-10-180-762-44	Sequence 44, Appl
29	220.5	13.5	281	15	US-10-241-258-44	Sequence 44, Appl
30	216	13.2	281	9	US-09-766-826-46	Sequence 46, Appl
31	215	13.1	281	9	US-09-765-205-24	Sequence 24, Appl
32	215	13.1	281	10	US-09-866-028-78	Sequence 78, Appl
33	215	13.1	281	10	US-09-944-449-78	Sequence 78, Appl
34	215	13.1	281	10	US-09-944-457-78	Sequence 78, Appl
35	215	13.1	281	10	US-09-944-862-78	Sequence 78, Appl
36	215	13.1	281	10	US-09-945-587-78	Sequence 78, Appl
37	215	13.1	281	10	US-09-945-015-78	Sequence 78, Appl
38	215	13.1	281	10	US-09-944-396-78	Sequence 78, Appl
39	215	13.1	281	10	US-09-944-097-78	Sequence 78, Appl
40	215	13.1	281	10	US-09-944-432-78	Sequence 78, Appl
41	215	13.1	281	10	US-09-943-762-78	Sequence 78, Appl
42	215	13.1	281	10	US-09-944-654-78	Sequence 78, Appl
43	215	13.1	281	10	US-09-943-851A-78	Sequence 78, Appl
44	215	13.1	281	10	US-09-944-413-78	Sequence 78, Appl
45	215	13.1	281	10	US-09-911-176B-2	Sequence 2, Appl

#### ALIGNMENTS

RESULT 1  
US-10-085-167-2  
; Sequence 2, Application US/10085167  
; Publication No. US20030170781A1  
; GENERAL INFORMATION:  
; APPLICANT: Holloway, James L.  
; TITLE OF INVENTION: SECRETED PROTEIN ZACRP4  
; FILE REFERENCE: 99-29  
; CURRENT APPLICATION NUMBER: US/10/085,167  
; CURRENT FILING DATE: 2002-02-27  
; PRIOR APPLICATION NUMBER: 60/141,928  
; PRIOR FILING DATE: 1999-07-01  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq For Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 329  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-085-167-2

Query Match	100.0%	Score 1639	DB 12	Length 329
Best Local Similarity	100.0%	Pred. No. 6.2e-149		
Matches 313	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	LGPTPGSSSLRKAFAAARTTPLGSEMAVTDXVYVNIIGSDPDTVGFRCPVPAY	60	
DB	17	LGPTPGSSSLRKAFAAARTTPLGSEMAVTDXVYVNIIGSDPDTVGFRCPVPAY	76	
QY	61	FFSFTAGAPKPKSLVNLVRNDEVQALAFDEQRPGARRAASOSAMQLDLYGDTWRL	120	
DB	77	FFSFTAGAPKPKSLVNLVRNDEVQALAFDEQRPGARRAASOSAMQLDLYGDTWRL	136	
QY	121	HGAPHYALGAPGATFGYLVADADADAPARGPAPPEPRSAFSAARTRSIVGSDAGCP	180	
DB	137	HGAPHYALGAPGATFGYLVADADADAPARGPAPPEPRSAFSAARTRSIVGSDAGCP	196	
QY	181	RHQLAPPTERTVNIIGSDPDAAGVFRCLPGAYFFSFTLGLPKRTISVTKMKRDEVQ	240	

Db 197 RHQPLAFDTEFVNIGDFFDAAGVFRRCRPGAYFFSFTLGKLPKRTLSVKLMKNRDEVOA 256  
Qy 241 MIYDDGASRRRMOQSVMALARGDVAWMLSHDHGAYSNHGKTYTTFSGFLVYPDLA 300  
Db 257 MIYDDGASRRRMOQSVMALARGDVAWMLSHDHGAYSNHGKTYTTFSGFLVYPDLA 316  
Qy 301 PAAPPLGASELL 313  
Db 317 PAAPPLGASELL 329

## RESULT 2

US-10-236-055A-14  
Sequence 14, Application US/10236055A  
Publication No. US20030134328A1  
GENERAL INFORMATION:  
APPLICANT: Basham, Beth E.  
APPLICANT: Forsythe, Ian  
APPLICANT: Gorman, Daniel M.  
APPLICANT: Matsson, Jeanine  
APPLICANT: Moshrefi, Mehdiad  
APPLICANT: Parham, Christi  
TITLE OF INVENTION: MAMMALIAN GENES; RELATED REAGENTS  
FILE REFERENCE: DX01343K  
CURRENT APPLICATION NUMBER: US/10/236,055A  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/317,988  
PRIOR FILING DATE: 2001-09-06  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 14  
LENGTH: 329  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-236-055A-14

Query Match 99.6%; Score 1633; DB 12; Length 329;  
Best Local Similarity 99.7%; Pred. No. 2,3e-148;  
Matches 312; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LGPTPGSSSELRSASFSAARTTPLEGTSEMAVTFDKYVNIIGDFFDVAATGQFRCRVPAY 60  
Db 17 LGPTPGSSSELRSASFSAARTTPLEGTSEMAVTFDKYVNIIGDFFDVAATGQFRCRVPAY 76  
Qy 61 FFSFTAGKAPKHSLSVMLVRNDEVQALAPDEQRPRGARRAASQSMQLDYGDTVMRL 120  
Db 77 FFSFTAGKAPKHSLSVMLVRNDEVQALAPDEQRPRGARRAASQSMQLDYGDTVMRL 136  
Qy 121 HGAPHYALGAPGATSGYLVAADADAPARGPPAPPEPRSAFSAARTSLVGSDDAGP 180  
Db 137 HGAPHYALGAPGATSGYLVAADADAPARGPPAPPEPRSAFSAARTSLVGSDDAGP 196  
Qy 181 RHQPLAFDTEFVNIGDFFDAAGVFRRCRPGAYFFSFTLGKLPKRTLSVKLMKNRDEVOA 240  
Db 197 RHQPLAFDTEFVNIGDFFDAAGVFRRCRPGAYFFSFTLGKLPKRTLSVKLMKNRDEVOA 256  
Qy 241 MIYDDGASRRRMOQSVMALARGDVAWMLSHDHGAYSNHGKTYTTFSGFLVYPDLA 300  
Db 257 MIYDDGASRRRMOQSVMALARGDVAWMLSHDHGAYSNHGKTYTTFSGFLVYPDLA 316  
Qy 301 PAAPPLGASELL 313  
Db 317 PAAPPLGASELL 329

## RESULT 3

US-10-236-055A-16  
Sequence 16, Application US/10236055A  
Publication No. US20030134328A1  
GENERAL INFORMATION:  
APPLICANT: Basham, Beth E.  
APPLICANT: Forsythe, Ian

APPLICANT: Gorman, Daniel M.  
APPLICANT: Matsson, Jeanine  
APPLICANT: Moshrefi, Mehdiad  
APPLICANT: Parham, Christi  
TITLE OF INVENTION: MAMMALIAN GENES; RELATED REAGENTS  
FILE REFERENCE: DX01343K  
CURRENT APPLICATION NUMBER: US/10/236,055A  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/317,988  
PRIOR FILING DATE: 2001-09-06  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 16  
LENGTH: 326  
TYPE: PRT  
ORGANISM: Mus musculus  
US-10-236-055A-16

Query Match 92.4%; Score 1514; DB 12; Length 326;  
Best Local Similarity 94.6%; Pred. No. 6.1e-137;  
Matches 295; Conservative 2; Mismatches 13; Indels 2; Gaps 2;

Qy 1 LGPTPGSSSELRSASFSAARTTPLEGTSEMAVTFDKYVNIIGDFFDVAATGQFRCRVPAY 60  
Db 17 LGPA-GPSSSELRSASFSAARTTPLEGTSEMAVTFDKYVNIIGDFFDVAATGQFRCRVPAY 75  
Qy 61 FFSFTAGKAPKHSLSVMLVRNDEVQALAPDEQRPRGARRAASQSMQLDYGDTVMRL 120  
Db 76 FFSFTAGKAPKHSLSVMLVRNDEVQALAPDEQRPRGARRAASQSMQLDYGDTVMRL 135  
Qy 121 HGAPHYALGAPGATSGYLVAADADAPARGPPAPPEPRSAFSAARTSLVGSDDAGP 180  
Db 136 HGAPHYALGAPGATSGYLVAADADAPARGPPAPPEPRSAFSAARTSLVGSDDAGP 194  
Qy 181 RHQPLAFDTEFVNIGDFFDAAGVFRRCRPGAYFFSFTLGKLPKRTLSVKLMKNRDEVOA 240  
Db 195 RHQPLAFDTEFVNIGDFFDAAGVFRRCRPGAYFFSFTLGKLPKRTLSVKLMKNRDEVOA 254  
Qy 241 MIYDDGASRRRMOQSVMALARGDVAWMLSHDHGAYSNHGKTYTTFSGFLVYPDLA 300  
Db 255 MIYDDGASRRRMOQSVMALARGDVAWMLSHDHGAYSNHGKTYTTFSGFLVYPDLA 314  
Qy 301 PAAPPLGASELL 312  
Db 315 AAGPPLKPEL 326

## RESULT 4

US-09-893-737-52  
Sequence 52, Application US/09893737  
Patent No. US20020110855A1  
GENERAL INFORMATION:  
APPLICANT: Sheppard, Paul O.  
APPLICANT: Preenell, Scott R.  
TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS  
FILE REFERENCE: 00-41  
CURRENT APPLICATION NUMBER: US/09/893,737  
CURRENT FILING DATE: 2001-06-28  
PRIOR APPLICATION NUMBER: US 60/215,446  
PRIOR FILING DATE: 2000-06-30  
NUMBER OF SEQ ID NOS: 329  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 52  
LENGTH: 252  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-893-737-52

Query Match 14.5%; Score 237.5; DB 10; Length 252;  
Best Local Similarity 43.0%; Pred. No. 1.4e-14;  
Matches 65; Conservative 12; Mismatches 61; Indels 13; Gaps 7;  
Qy 152 GPPAPPEPR-SAFSAARTSLVGSDDAGP RHQPLAFDTEFVNIGDFFDAAGVFRRCRL 209

Db 106 GPRGACRAVAFVSGRREGHSD-----HFQAVPFTLVLNDGAFDLAAGFLCTV 160  
Qy 210 PGAYFSTLGLPKRTLSVKLMKNDVQAMITYDGSRRKQSQSVMLALRGDAVW 269  
Db 161 PGVYFLSLVHVHTWNYKETYLHMLNR-RPAVALVYQ-PSERSVMOQOSIMLLIAGDAVW 218  
Qy 270 --LASHDHGCVAYSNHGK-YTFSGFLVYP 297  
Db 219 VAMFORDRDN-AIYGBHGDLVTFSGHLVXP 248

RESULT 5  
US-10-091-458-43  
; Sequence 43, Application US/10091458  
; Publication No. US2003008627A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT206CJ  
; CURRENT APPLICATION NUMBER: US/10/091,458  
; CURRENT FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 09/764,900  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/214,886  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 60/217,487  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,758  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,963  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/217,496  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,447  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/218,290  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/225,757  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/226,868  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/216,647  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: 60/225,267  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/216,880  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: 60/225,270  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/251,869  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/235,834  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/234,274  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: 60/234,223  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: 60/228,924  
; PRIOR FILING DATE: 2000-08-30  
; PRIOR APPLICATION NUMBER: 60/224,518  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/236,369  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/224,519  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,964  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/241,809

; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/249,239  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/236,327  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/241,785  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/244,617  
; PRIOR FILING DATE: 2000-11-01  
; PRIOR APPLICATION NUMBER: 60/225,268  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/236,368  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/251,856  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/251,868  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/229,344  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/234,997  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: 60/229,343  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,345  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,287  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,513  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: 60/231,413  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/229,509  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: 60/236,367  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/237,039  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,038  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/236,370  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/236,802  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,037  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,040  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/240,960  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/239,935  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: 60/239,937  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: 60/241,787  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/246,474  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/246,532  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/249,216  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,210  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/226,681  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/225,759  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/225,213  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/227,182  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/225,214

PRIOR APPLICATION NUMBER: 60/235, 836  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: 60/230, 438  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/215, 135  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: 60/225, 266  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/249, 218  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249, 208  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249, 213  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249, 212  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249, 207  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249, 245  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249, 244  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249, 217  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249, 211  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249, 215  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249, 264  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249, 214  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249, 297  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/232, 400  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/231, 242  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/232, 081  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/232, 080  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/231, 414  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/231, 244  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/233, 064  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/233, 063  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232, 397  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232, 399  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232, 401  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/241, 808  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241, 826  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241, 786  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241, 221  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/246, 475  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/231, 243  
PRIOR FILING DATE: 2000-09-08

Query Match 14.0%; Score 230; DB 15; Length 202;  
Best Local Similarity 38.4%; Pred. No. 5, 5e-14;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 147 DAPARGP-----PAPPEPSAFSAARTSLVSGDAGCPRHQPLAFDTEFNIG 196  
DB 49 DVPVTPAATILPHVAVYLPQGMRAVFAAART-----SNLAPGTLDDQIVFDLLNNIG 103  
QY 197 DFDAAAGVRCRLPGAYFFSFTLGKLP-RKTLVYKLMKNRDEVOQMIYDDGASRRRQOS 255  
DB 104 TFDLQLGFRNCFVNGTYVFIHMLKLVAVNPLYVLMKNEEVLVSAYVANDGAP-DHETAS 162  
QY 256 QSWMLALRRGDVAWMLSHDHGXYGA-YSNHGKTYTFSGFLVYPD 298  
DB 163 NHAILQLFQGDQIWLRLH-----RGAIYGSWMKSTFSGYLLYOD 202

RESULT 6  
US-10-236-055A-2  
Sequence 2, Application US/10236055A  
Publication No. US20030134528A1  
GENERAL INFORMATION:  
APPLICANT: Basham, Beth E.  
APPLICANT: Forsythe, Ian  
APPLICANT: Gorman, Daniel M.  
APPLICANT: Mattson, Jeanine  
APPLICANT: Moshrefi, Mehرداد  
APPLICANT: Parham, Christi  
TITLE OF INVENTION: MAMMALIAN GENES, RELATED REAGENTS  
FILE REFERENCE: DX01343X  
CURRENT APPLICATION NUMBER: US/10/236, 055A  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/317, 988  
PRIOR FILING DATE: 2001-09-06  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 2  
LENGTH: 420  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-236-055A-2

Query Match 14.0%; Score 230; DB 12; Length 420;  
Best Local Similarity 38.4%; Pred. No. 1, 5e-13;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 147 DAPARGP-----PAPPEPSAFSAARTSLVSGDAGCPRHQPLAFDTEFNIG 196  
DB 267 DVPVTPAATILPHVAVYLPQGMRAVFAAART-----SNLAPGTLDDQIVFDLLNNIG 321  
QY 197 DFDAAAGVRCRLPGAYFFSFTLGKLP-RKTLVYKLMKNRDEVOQMIYDDGASRRRQOS 255  
DB 322 TFDLQLGFRNCFVNGTYVFIHMLKLVAVNPLYVLMKNEEVLVSAYVANDGAP-DHETAS 380  
QY 256 QSWMLALRRGDVAWMLSHDHGXYGA-YSNHGKTYTFSGFLVYPD 298  
DB 381 NHAILQLFQGDQIWLRLH-----RGAIYGSWMKSTFSGYLLYOD 420

RESULT 7  
US-09-738-973-185  
Sequence 185, Application US/09738973  
Patent No. US20020110563A1  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Henderson, Robert A.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Fling, Steven P.  
APPLICANT: Mohamath, Raodoh  
APPLICANT: Algate, Paul A.  
APPLICANT: Secrist, Heather  
APPLICANT: Indirias, Carol Yoseph  
APPLICANT: Benson, Darin R.  
APPLICANT: Elliot, Mark  
APPLICANT: Mannion, Jane  
APPLICANT: Kalos, Michael D.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

;; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
;; FILE REFERENCE: 210121.475C9  
;; CURRENT APPLICATION NUMBER: US/09/738.973  
;; CURRENT FILING DATE: 2000-12-14  
;; NUMBER OF SEQ ID NOS: 587  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 185  
;; LENGTH: 746  
;; TYPE: PRT  
;; ORGANISM: Homo sapien  
US-09-738-973-185

Query Match 14.0%; Score 230; DB 10; Length 746;  
Best Local Similarity 38.4%; Pred. No. 3.2e-13;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 147 DAPARGP-----PAPPEPSAFSAARTSLVSGDAGCPGRHOPLAFTFEFVNIIG 196  
DB 593 DVPVTNPATILPVAHYVPLPQOMRVAFAFSART-----SNLAPGTLDPQIVFDLLNNLIGE 647  
QY 197 DFDAAAGVFRCLPGAYFFSFTLGKLP-RKTLVYLMKNRDEVQAMTYDDGASRRREMOS 255  
DB 648 TFDLQGRFNCVPNGVTYVFIHMLKLAIVNPLYVNLKMBEVLVSAYANDGAP-DHETAS 706  
QY 256 QSVMLALRRGDAVWLSSHHDGYGA-YSNHGKTYITFGSLVLYPD 298  
DB 707 NHAILQLFGDQIWLRLH---RGAIVGSSWKYSTFGSLYLYPD 746

RESULT 8  
US-09-854-133-185  
; Sequence 185, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Henderson, Radoch  
; APPLICANT: Benson, Robert A.  
; APPLICANT: Benson, Darin R.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854.133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 185  
; LENGTH: 746  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-09-854-133-185

Query Match 14.0%; Score 230; DB 10; Length 746;  
Best Local Similarity 38.4%; Pred. No. 3.2e-13;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 147 DAPARGP-----PAPPEPSAFSAARTSLVSGDAGCPGRHOPLAFTFEFVNIIG 196  
DB 593 DVPVTNPATILPVAHYVPLPQOMRVAFAFSART-----SNLAPGTLDPQIVFDLLNNLIGE 647  
QY 197 DFDAAAGVFRCLPGAYFFSFTLGKLP-RKTLVYLMKNRDEVQAMTYDDGASRRREMOS 255  
DB 648 TFDLQGRFNCVPNGVTYVFIHMLKLAIVNPLYVNLKMBEVLVSAYANDGAP-DHETAS 706  
QY 256 QSVMLALRRGDAVWLSSHHDGYGA-YSNHGKTYITFGSLVLYPD 298  
DB 707 NHAILQLFGDQIWLRLH---RGAIVGSSWKYSTFGSLYLYPD 746

RESULT 9  
US-10-144-649A-185  
; Sequence 185, Application US/10144649A  
; Publication No. US20030118599A1

;; GENERAL INFORMATION:  
;; APPLICANT: Lodes, Michael J.  
;; APPLICANT: Wang, Tonglong  
;; APPLICANT: Fan, Liqun  
;; APPLICANT: Algate, Paul A.  
;; APPLICANT: McNeill, Patricia D.  
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
;; FILE REFERENCE: 210121.475C11  
;; CURRENT APPLICATION NUMBER: US/10/144.649A  
;; CURRENT FILING DATE: 2002-08-21  
;; NUMBER OF SEQ ID NOS: 749  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 185  
;; LENGTH: 746  
;; TYPE: PRT  
;; ORGANISM: Homo sapien  
US-10-144-649A-185

Query Match 14.0%; Score 230; DB 15; Length 746;  
Best Local Similarity 38.4%; Pred. No. 3.2e-13;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 147 DAPARGP-----PAPPEPSAFSAARTSLVSGDAGCPGRHOPLAFTFEFVNIIG 196  
DB 593 DVPVTNPATILPVAHYVPLPQOMRVAFAFSART-----SNLAPGTLDPQIVFDLLNNLIGE 647  
QY 197 DFDAAAGVFRCLPGAYFFSFTLGKLP-RKTLVYLMKNRDEVQAMTYDDGASRRREMOS 255  
DB 648 TFDLQGRFNCVPNGVTYVFIHMLKLAIVNPLYVNLKMBEVLVSAYANDGAP-DHETAS 706  
QY 256 QSVMLALRRGDAVWLSSHHDGYGA-YSNHGKTYITFGSLVLYPD 298  
DB 707 NHAILQLFGDQIWLRLH---RGAIVGSSWKYSTFGSLYLYPD 746

RESULT 10  
US-10-309-422-10  
; Sequence 10, Application US/10309422  
; Publication No. US20030139587A1  
; GENERAL INFORMATION:  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wilganowski, Nachaniel L.  
; APPLICANT: Turner, C. Alexander Jr.  
; TITLE OF INVENTION: No. US20030139587A1 Human Proteins and Polynucleotides Encoding  
; FILE REFERENCE: LEX-0142-USA  
; CURRENT APPLICATION NUMBER: US/10/309.422  
; CURRENT FILING DATE: 2002-12-03  
; PRIOR APPLICATION NUMBER: US/09/798.771  
; PRIOR FILING DATE: 2001-03-02  
; PRIOR APPLICATION NUMBER: US 60/186,557  
; PRIOR FILING DATE: 2000-03-02  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 908  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-309-422-10

Query Match 14.0%; Score 230; DB 12; Length 908;  
Best Local Similarity 38.4%; Pred. No. 4.2e-13;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 147 DAPARGP-----PAPPEPSAFSAARTSLVSGDAGCPGRHOPLAFTFEFVNIIG 196  
DB 755 DVPVTNPATILPVAHYVPLPQOMRVAFAFSART-----SNLAPGTLDPQIVFDLLNNLIGE 809  
QY 197 DFDAAAGVFRCLPGAYFFSFTLGKLP-RKTLVYLMKNRDEVQAMTYDDGASRRREMOS 255  
DB 810 TFDLQGRFNCVPNGVTYVFIHMLKLAIVNPLYVNLKMBEVLVSAYANDGAP-DHETAS 868  
QY 256 QSVMLALRRGDAVWLSSHHDGYGA-YSNHGKTYITFGSLVLYPD 298

Db 869 NHAILQLFGDQIWLRLH-----RGAIVGSSWKYSTFGYLLYQD 908

RESULT 11

US-10-309-422-22

Sequence 22, Application US/10309422

Publication No. US20030139587A1

GENERAL INFORMATION:

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

APPLICANT: Turner, C. Alexander Jr.

TITLE OF INVENTION: No. US20030139587A1el Human Proteins and Polynucleotides Encoding

FILE REFERENCE: LEX-0142-USA

CURRENT APPLICATION NUMBER: US/10/309,422

CURRENT FILING DATE: 2002-12-03

PRIOR APPLICATION NUMBER: US/09/798,771

PRIOR FILING DATE: 2001-03-02

PRIOR APPLICATION NUMBER: US 60/186,557

PRIOR FILING DATE: 2000-03-02

NUMBER OF SEQ ID NOS: 43

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 22

LENGTH: 909

TYPE: PRT

ORGANISM: homo sapiens

US-10-309-422-22

Query Match 14.0%; Score 230; DB 12; Length 909;

Best Local Similarity 38.4%; Pred. No. 4.5e-13;

Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

Qy 147 DAPARGP-----PAPPEPSAFSAARTSLVSGDAGCPRHQPLAFDTEFVNI 196

Db 756 DVPVTNPATITLPHVYVLPQGMVAFSART-----SNLAGTLDQPIVDFLLNNGE 810

Qy 197 DPDAAGVRCRLPGAYFFSFTLGKLP-RKTLVYKLMKNRDEVQAMIVDDGASRRRMO 255

Db 811 TFDQLGRFNCVPNGTYVFIHMLKLVAVNPVLYNLMKNBEVLVSAYANDGAP-DHETAS 869

Qy 256 QSWMLALRRGDAVWLLSHDHGYGGA-YSNHGKTYTFSGFLVYPD 298

Db 870 NHAILQLFGDQIWLRLH-----RGAIVGSSWKYSTFGYLLYQD 909

RESULT 12

US-10-309-422-14

Sequence 14, Application US/10309422

Publication No. US20030139587A1

GENERAL INFORMATION:

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

APPLICANT: Turner, C. Alexander Jr.

TITLE OF INVENTION: No. US20030139587A1el Human Proteins and Polynucleotides Encoding

FILE REFERENCE: LEX-0142-USA

CURRENT APPLICATION NUMBER: US/10/309,422

CURRENT FILING DATE: 2002-12-03

PRIOR APPLICATION NUMBER: US/09/798,771

PRIOR FILING DATE: 2001-03-02

PRIOR APPLICATION NUMBER: US 60/186,557

PRIOR FILING DATE: 2000-03-02

NUMBER OF SEQ ID NOS: 43

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 14

LENGTH: 957

TYPE: PRT

ORGANISM: homo sapiens

US-10-309-422-14

Query Match 14.0%; Score 230; DB 12; Length 957;

Best Local Similarity 38.4%; Pred. No. 4.5e-13;

Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

Qy 147 DAPARGP-----PAPPEPSAFSAARTSLVSGDAGCPRHQPLAFDTEFVNI 196

Db 804 DVPVTNPATITLPHVYVLPQGMVAFSART-----SNLAGTLDQPIVDFLLNNGE 858

Qy 197 DPDAAGVRCRLPGAYFFSFTLGKLP-RKTLVYKLMKNRDEVQAMIVDDGASRRRMO 255

Db 859 TFDQLGRFNCVPNGTYVFIHMLKLVAVNPVLYNLMKNBEVLVSAYANDGAP-DHETAS 917

Qy 256 QSWMLALRRGDAVWLLSHDHGYGGA-YSNHGKTYTFSGFLVYPD 298

Db 918 NHAILQLFGDQIWLRLH-----RGAIVGSSWKYSTFGYLLYQD 957

RESULT 13

US-10-309-422-26

Sequence 26, Application US/10309422

Publication No. US20030139587A1

GENERAL INFORMATION:

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

APPLICANT: Turner, C. Alexander Jr.

TITLE OF INVENTION: No. US20030139587A1el Human Proteins and Polynucleotides Encoding

FILE REFERENCE: LEX-0142-USA

CURRENT APPLICATION NUMBER: US/10/309,422

CURRENT FILING DATE: 2002-12-03

PRIOR APPLICATION NUMBER: US/09/798,771

PRIOR FILING DATE: 2001-03-02

PRIOR APPLICATION NUMBER: US 60/186,557

PRIOR FILING DATE: 2000-03-02

NUMBER OF SEQ ID NOS: 43

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 26

LENGTH: 958

TYPE: PRT

ORGANISM: homo sapiens

US-10-309-422-26

Query Match 14.0%; Score 230; DB 12; Length 958;

Best Local Similarity 38.4%; Pred. No. 4.5e-13;

Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

Qy 147 DAPARGP-----PAPPEPSAFSAARTSLVSGDAGCPRHQPLAFDTEFVNI 196

Db 805 DVPVTNPATITLPHVYVLPQGMVAFSART-----SNLAGTLDQPIVDFLLNNGE 859

Qy 197 DPDAAGVRCRLPGAYFFSFTLGKLP-RKTLVYKLMKNRDEVQAMIVDDGASRRRMO 255

Db 860 TFDQLGRFNCVPNGTYVFIHMLKLVAVNPVLYNLMKNBEVLVSAYANDGAP-DHETAS 918

Qy 256 QSWMLALRRGDAVWLLSHDHGYGGA-YSNHGKTYTFSGFLVYPD 298

Db 919 NHAILQLFGDQIWLRLH-----RGAIVGSSWKYSTFGYLLYQD 958

RESULT 14

US-10-309-422-18

Sequence 18, Application US/10309422

Publication No. US20030139587A1

GENERAL INFORMATION:

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

APPLICANT: Turner, C. Alexander Jr.

TITLE OF INVENTION: No. US20030139587A1el Human Proteins and Polynucleotides Encoding

FILE REFERENCE: LEX-0142-USA

CURRENT APPLICATION NUMBER: US/10/309,422

CURRENT FILING DATE: 2002-12-03

PRIOR APPLICATION NUMBER: US/09/798,771

PRIOR FILING DATE: 2001-03-02

PRIOR APPLICATION NUMBER: US 60/186,557

PRIOR FILING DATE: 2000-03-02

NUMBER OF SEQ ID NOS: 43

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 18

LENGTH: 992  
TYPE: PRT  
ORGANISM: homo sapiens  
US-10-309-422-18

Query Match 14.0%; Score 230; DB 12; Length 992;  
Best Local Similarity 38.4%; Pred. No. 4,7e-13;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 147 DAPARGP-----PAPPEPRSAFSAARTSLVSGDAGPGRHOPLAFDTEFVNIG 196  
DB 839 DVPVTNPATILPVHVYPLPQMRVAFSAART-----SNLAPGTLDPFVFDLLNINIG 893  
QY 197 DEDAAAGVFRCLLPAGYFFSFTLGKLP--RKTLSVTKMKRDEVOAMTYDDGASRRREMOS 255  
DB 894 TFDLQIGRNCPCVNGTYVFI FHLMLKLA VNPPLYVNLKNEEVLVSAYANDGAP-DHETAS 952  
QY 256 QSVMLALRRGDVWLLSHDHG YGA-YSNHGKYITFSGFLVYPD 298  
DB 953 NHAILQLFGDDIWLRLH---RGAITYGSSWKSTFSGTLYOD 992

## RESULT 15

US-10-309-422-30  
Sequence 30, Application US/10309422  
Publication No. US20030139587A1  
GENERAL INFORMATION:  
APPLICANT: Walke, D. Wade  
APPLICANT: Wilganowski, Nathaniel L.  
TITLE OF INVENTION: No. US20030139587A1el Human Proteins and Polynucleotides Encoding  
FILE REFERENCE: LEX-0142-USA  
CURRENT APPLICATION NUMBER: US/10/309,422  
CURRENT FILING DATE: 2002-12-03  
PRIOR FILING DATE: 2001-03-02  
PRIOR APPLICATION NUMBER: US/09/798,771  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 60/186,557  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 30  
LENGTH: 993  
TYPE: PRT  
ORGANISM: homo sapiens  
US-10-309-422-30

Query Match 14.0%; Score 230; DB 12; Length 993;  
Best Local Similarity 38.4%; Pred. No. 4,7e-13;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 147 DAPARGP-----PAPPEPRSAFSAARTSLVSGDAGPGRHOPLAFDTEFVNIG 196  
DB 840 DVPVTNPATILPVHVYPLPQMRVAFSAART-----SNLAPGTLDPFVFDLLNINIG 894  
QY 197 DEDAAAGVFRCLLPAGYFFSFTLGKLP--RKTLSVTKMKRDEVOAMTYDDGASRRREMOS 255  
DB 895 TFDLQIGRNCPCVNGTYVFI FHLMLKLA VNPPLYVNLKNEEVLVSAYANDGAP-DHETAS 953  
QY 256 QSVMLALRRGDVWLLSHDHG YGA-YSNHGKYITFSGFLVYPD 298  
DB 954 NHAILQLFGDDIWLRLH---RGAITYGSSWKSTFSGTLYOD 993

Search completed: January 12, 2004, 08:33:36  
Job time : 39.1098 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:14 : Search time 15.911 Seconds  
(without alignments)  
831.284 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_17\_329  
Perfect score: 1639  
Sequence: 1 LGPTGPGSSLSRFAFSAAR.....LVYPLAPAPPGGLASSEL 313

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	230	14.0	746	4	US-09-370-838-185 Sequence 185, App
2	220.5	13.5	281	3	US-09-118-408-44 Sequence 44, App1
3	220.5	13.5	281	4	US-09-506-855-44 Sequence 44, App1
4	220.5	13.5	281	4	US-09-911-176B-44 Sequence 44, App1
5	220.5	13.5	281	4	US-09-619-740-44 Sequence 44, App1
6	220.5	13.5	281	4	US-09-506-855-44 Sequence 44, App1
7	215	13.1	281	3	US-09-118-408-2 Sequence 2, App1
8	215	13.1	281	4	US-09-506-855-2 Sequence 2, App1
9	215	13.1	281	4	US-09-911-176B-2 Sequence 2, App1
10	215	13.1	281	4	US-09-619-740-2 Sequence 2, App1
11	215	13.1	281	4	US-09-506-852-2 Sequence 2, App1
12	213	13.0	228	4	US-09-336-536-4 Sequence 4, App1
13	213	13.0	243	3	US-09-140-804-2 Sequence 3, App1
14	213	13.0	243	4	US-09-336-536-3 Sequence 3, App1
15	213	13.0	243	4	US-09-686-838B-2 Sequence 2, App1
16	207	12.6	243	3	US-09-188-930-295 Sequence 295, App
17	207	12.6	243	4	US-09-312-283C-295 Sequence 11, App1
18	202	12.3	228	4	US-09-336-536-11 Sequence 11, App1
19	202	12.3	259	4	US-09-336-536-10 Sequence 10, App1
20	200	12.2	294	3	US-09-996-243-47 Sequence 47, App1
21	200	12.2	294	4	US-09-188-930-294 Sequence 294, App
22	200	12.2	294	4	US-09-312-283C-294 Sequence 294, App
23	197	12.0	285	4	US-09-312-283C-382 Sequence 382, App
24	196	12.0	260	4	US-09-489-847-198 Sequence 198, App
25	196	12.0	287	4	US-09-489-847-198 Sequence 349, App
26	192	11.7	231	4	US-09-530-423-2 Sequence 2, App1
27	192	11.7	244	2	US-08-463-911-7 Sequence 7, App1

28	192	11.7	244	3	US-09-140-804-3 Sequence 3, App1
29	192	11.7	244	4	US-09-336-536-20 Sequence 20, App1
30	192	11.7	244	4	US-09-530-423-1 Sequence 1, App1
31	192	11.7	244	4	US-09-686-838B-3 Sequence 3, App1
32	192	11.7	244	4	US-09-911-176B-48 Sequence 48, App1
33	192	11.7	244	4	US-09-552-225A-3 Sequence 3, App1
34	192	11.7	244	4	US-09-619-740-51 Sequence 51, App1
35	192	11.7	244	4	US-09-776-976-6 Sequence 6, App1
36	192	11.7	244	4	US-09-909-547-6 Sequence 6, App1
37	192	11.7	244	4	US-09-569-852B-6 Sequence 6, App1
38	188	11.5	247	2	US-08-463-911-2 Sequence 2, App1
39	188	11.5	247	4	US-09-776-976-4 Sequence 4, App1
40	188	11.5	247	4	US-09-909-547-4 Sequence 4, App1
41	187	11.4	247	3	US-09-140-804-8 Sequence 8, App1
42	187	11.4	247	3	US-09-118-408-3 Sequence 3, App1
43	187	11.4	247	4	US-09-506-855-3 Sequence 3, App1
44	187	11.4	247	4	US-09-686-838B-8 Sequence 8, App1
45	187	11.4	247	4	US-09-911-176B-3 Sequence 3, App1

## ALIGNMENTS

```

RESULT 1
US-09-370-838-185
; Sequence 185, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 185
; LENGTH: 746
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-370-838-185
Query Match          14.0% Score 230; DB 4; Length 746;
Best Local Similarity 38.4% Pred. No. 7.8e-15;
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 147 DAPARGP-----DAPPEPRSAFSAARTSLVSGDAGGPRHQPAPDTEFVNIG 196
DB 593 DVPVTNPATILPVHYVPLPOQMRVAFSAART-----SNLAPGTLDQIVFDLLNLTNGE 647
QY 197 DPDAAGGFRRLRGAYFFSTLGLKP-RTKLSVYLMKNRBEVQAMITDDASRRRERG 255
DB 648 TFDLQGFNCPVNGTYVIFHMLKLAVNPVLYVLMNBEVLVSAYVNDGAP-DHETAS 706
QY 256 QSVMLALRGDAVWMLSHDHGYGA-YSNHGKYITFGSFLVYPD 298
DB 707 NHAILQLFQSDQIMLRH-----RGAIYSSMKYITFGSFLVLYQD 746

RESULT 2
US-09-118-408-44
; Sequence 44, Application US/09118408A
; Patent No. 6265544
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-30
; CURRENT APPLICATION NUMBER: US/09/118,408A

```

;; CURRENT FILING DATE: 1998-07-17  
;; EARLIER APPLICATION NUMBER: 60/053,154  
;; EARLIER FILING DATE: 1997-07-18  
;; NUMBER OF SEQ ID NOS: 47  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 44  
;; LENGTH: 281  
;; TYPE: PRT  
;; ORGANISM: Mus musculus  
US-09-118-408-44

Query Match 13.5%; Score 220.5; DB 3; Length 281;  
Best Local Similarity 39.2%; Pred. No. 1.8e-14;  
Matches 56; Conservative 20; Mismatches 58; Indels 9; Gaps 4;

QY 161 SAFSAARTSLVGDAGPGRPHOPLAFDTEFVNIGDPPDAAGVPRCLPGAYFFSFTLG 220  
DB 146 AAFSVGRKKALHSND-----YFQPVVFDTEFVNLYKHFMFTGKFCYVPGIYFFSLNVH 200  
QY 221 KLPRKTLVSKMKNRDEVQAMITYDDGASRRREMOSQSVMLALRRGDVW--LLSHDHGY 278  
DB 201 TNNQKETYLIHKMNEEV-VILYAO-VSDRSIMOSQSLMMLREBDEVWVRLFKGERENA 258  
QY 279 GAYSNHGKYYTFSGLVYPPDLAP 301  
DB 259 IFSDEFDTYITFSGYLVKPASEP 281

RESULT 3  
US-09-506-855-44  
; Sequence 44, Application US/09506855  
; Patent No. 6448221  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; APPLICANT: Lasser, Gerald W.  
; APPLICANT: Bishop, Paul D.  
; TITLE OF INVENTION: INHIBITORS FOR USE IN HEMOSTASIS AND  
; FILE REFERENCE: 99-12  
; CURRENT APPLICATION NUMBER: US/09/506,855  
; CURRENT FILING DATE: 2000-02-17  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 44  
; LENGTH: 281  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-506-855-44

Query Match 13.5%; Score 220.5; DB 4; Length 281;  
Best Local Similarity 39.2%; Pred. No. 1.8e-14;  
Matches 56; Conservative 20; Mismatches 58; Indels 9; Gaps 4;

QY 161 SAFSAARTSLVGDAGPGRPHOPLAFDTEFVNIGDPPDAAGVPRCLPGAYFFSFTLG 220  
DB 146 AAFSVGRKKALHSND-----YFQPVVFDTEFVNLYKHFMFTGKFCYVPGIYFFSLNVH 200  
QY 221 KLPRKTLVSKMKNRDEVQAMITYDDGASRRREMOSQSVMLALRRGDVW--LLSHDHGY 278  
DB 201 TNNQKETYLIHKMNEEV-VILYAO-VSDRSIMOSQSLMMLREBDEVWVRLFKGERENA 258  
QY 279 GAYSNHGKYYTFSGLVYPPDLAP 301  
DB 259 IFSDEFDTYITFSGYLVKPASEP 281

RESULT 4  
US-09-911-176B-44  
; Sequence 44, Application US/0991176B  
; Patent No. 6518403  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; TITLE OF INVENTION: ANTIBODIES THAT BIND AN

;; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOG  
;; FILE REFERENCE: 97-30D1  
;; CURRENT APPLICATION NUMBER: US/09/911,176B  
;; CURRENT FILING DATE: 2001-07-23  
;; PRIOR APPLICATION NUMBER: 09/118,408  
;; PRIOR FILING DATE: 1998-07-17  
;; PRIOR APPLICATION NUMBER: 60/053,154  
;; PRIOR FILING DATE: 1997-07-18  
;; NUMBER OF SEQ ID NOS: 52  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 44  
;; LENGTH: 281  
;; TYPE: PRT  
;; ORGANISM: Mus musculus  
US-09-911-176B-44

Query Match 13.5%; Score 220.5; DB 4; Length 281;  
Best Local Similarity 39.2%; Pred. No. 1.8e-14;  
Matches 56; Conservative 20; Mismatches 58; Indels 9; Gaps 4;

QY 161 SAFSAARTSLVGDAGPGRPHOPLAFDTEFVNIGDPPDAAGVPRCLPGAYFFSFTLG 220  
DB 146 AAFSVGRKKALHSND-----YFQPVVFDTEFVNLYKHFMFTGKFCYVPGIYFFSLNVH 200  
QY 221 KLPRKTLVSKMKNRDEVQAMITYDDGASRRREMOSQSVMLALRRGDVW--LLSHDHGY 278  
DB 201 TNNQKETYLIHKMNEEV-VILYAO-VSDRSIMOSQSLMMLREBDEVWVRLFKGERENA 258  
QY 279 GAYSNHGKYYTFSGLVYPPDLAP 301  
DB 259 IFSDEFDTYITFSGYLVKPASEP 281

RESULT 5  
US-09-619-740-44  
; Sequence 44, Application US/09619740  
; Patent No. 6544946  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; APPLICANT: Lasser, Gerald W.  
; APPLICANT: Bishop, Paul D.  
; TITLE OF INVENTION: INHIBITORS FOR USE IN HEMOSTASIS AND IMMUNE FUNCTION  
; FILE REFERENCE: 99-12C3  
; CURRENT APPLICATION NUMBER: US/09/619,740  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/253,604  
; PRIOR FILING DATE: 1999-02-19  
; PRIOR APPLICATION NUMBER: 09/444,794  
; PRIOR FILING DATE: 1999-11-22  
; PRIOR APPLICATION NUMBER: 09/506,855  
; PRIOR FILING DATE: 2000-02-17  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 44  
; LENGTH: 281  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-619-740-44

Query Match 13.5%; Score 220.5; DB 4; Length 281;  
Best Local Similarity 39.2%; Pred. No. 1.8e-14;  
Matches 56; Conservative 20; Mismatches 58; Indels 9; Gaps 4;

QY 161 SAFSAARTSLVGDAGPGRPHOPLAFDTEFVNIGDPPDAAGVPRCLPGAYFFSFTLG 220  
DB 146 AAFSVGRKKALHSND-----YFQPVVFDTEFVNLYKHFMFTGKFCYVPGIYFFSLNVH 200  
QY 221 KLPRKTLVSKMKNRDEVQAMITYDDGASRRREMOSQSVMLALRRGDVW--LLSHDHGY 278  
DB 201 TNNQKETYLIHKMNEEV-VILYAO-VSDRSIMOSQSLMMLREBDEVWVRLFKGERENA 258  
QY 279 GAYSNHGKYYTFSGLVYPPDLAP 301

Db 259 IFSDEFTYITFSGLVVKPASEP 281

RESULT 6

US-09-506-852-44

Sequence 44, Application US/09506852

Patent No. 6566499

GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.

TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS

FILE REFERENCE: 97-30

CURRENT APPLICATION NUMBER: US/09/506,852

CURRENT FILING DATE: 2000-02-17

EARLIER APPLICATION NUMBER: 60/053,154

EARLIER FILING DATE: 1997-07-18

NUMBER OF SEQ ID NOS: 44

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 44

LENGTH: 281

TYPE: PRT

ORGANISM: Mus musculus

US-09-506-852-44

Query Match 13.5%; Score 220.5; DB 4; Length 281;

Best Local Similarity 39.2%; Pred. No. 1.8e-14;

Matches 56; Conservative 20; Mismatches 58; Indels 9; Gaps 4;

Db 161 SAFSAARTSLVSDAGPGPRHOPLAFTDFEVNIGDPPDAAGVRCRLPGAYFFSFTLG 220

146 AAFSVGRKKALHSND----YFQPVVFDTEFVNLYKHFMFTGKFCYCVPGIYFFSLNVH 200

Qy 221 KLPRKTLVSKMKRDEVOAMTYDDGASRRREMOSQSVLALRRGDVW--LLSDHDGY 278

Db 201 TWNOKETYLHIMKNEEV-VIIYAO-VSDRSIMOSQSLMELREDEWVWVLFKGERENA 258

Qy 279 GAYSNHGKIYTFSGFLVVPDLAP 301

Db 259 IFSDEFTYITFSGLVVKPASEP 281

RESULT 7

US-09-118-408-2

Sequence 2, Application US/09118408A

Patent No. 6265544

GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.

TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS

FILE REFERENCE: 97-30

CURRENT APPLICATION NUMBER: US/09/118,408A

CURRENT FILING DATE: 1998-07-17

EARLIER APPLICATION NUMBER: 60/053,154

EARLIER FILING DATE: 1997-07-18

NUMBER OF SEQ ID NOS: 47

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 2

LENGTH: 281

TYPE: PRT

ORGANISM: Homo sapien

US-09-118-408-2

Query Match 13.1%; Score 215; DB 3; Length 281;

Best Local Similarity 30.2%; Pred. No. 6.6e-14;

Matches 77; Conservative 26; Mismatches 86; Indels 66; Gaps 11;

Db 53 RCRVGAFFSFYTAGKAPKSLVWLVNRDEVOALAFDEGRPGARRAASQSMLOLDY 112

75 RCCDPTGSMYPATA--VQINITITLKGKGR-----GDRG-----LOGKY 113

Qy 113 GDT--VWLRLLHGAAP--HYALGAPGATFSGYLIVYADADADAPARGPAPPEPRSAFSAAR 167

Db 114 GKTGSAGARGHTGPKGQKSGWAPGRCKSH--YA-----AFSVGR 152

Qy 168 TRSLVGSADGPGPRH-----QPLAFDTEFVNIGDPPDAAGVRCRLPGAYFFSFTLGKL 222

Db 153 KK-----PMSNHYQTVIFDTEFVNLYDHFNNFTGKFCYCVPGLYFFSLNVHTW 202

Qy 223 PRKTLVSKMKRDEVOAMTYDDGASRRREMOSQSVLALRRGDVWVLLSDHDGYGAYS 282

Db 203 NOKETYLHIMKNEEVVILFAQVG--DRSIMOSQSLMELREDOQVWVRLYKGERENAIF 260

Qy 283 NH--GKYTFSGFLY 295

Db 261 SEELDTYITFSGLY 275

RESULT 8

US-09-506-855-2

Sequence 2, Application US/09506855

Patent No. 6448221

GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.

APPLICANT: Bishop, Paul D.

TITLE OF INVENTION: INHIBITORS FOR USE IN HEMOSTASIS AND

FILE REFERENCE: 99-12

CURRENT APPLICATION NUMBER: US/09/506,855

CURRENT FILING DATE: 2000-02-17

NUMBER OF SEQ ID NOS: 50

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 2

LENGTH: 281

TYPE: PRT

ORGANISM: Homo sapien

US-09-506-855-2

Query Match 13.1%; Score 215; DB 4; Length 281;

Best Local Similarity 30.2%; Pred. No. 6.6e-14;

Matches 77; Conservative 26; Mismatches 86; Indels 66; Gaps 11;

Qy 53 RCRVGAFFSFYTAGKAPKSLVWLVNRDEVOALAFDEGRPGARRAASQSMLOLDY 112

Db 75 RCCDPTGSMYPATA--VQINITITLKGKGR-----GDRG-----LOGKY 113

Qy 113 GDT--VWLRLLHGAAP--HYALGAPGATFSGYLIVYADADADAPARGPAPPEPRSAFSAAR 167

Db 114 GKTGSAGARGHTGPKGQKSGWAPGRCKSH--YA-----AFSVGR 152

Qy 168 TRSLVGSADGPGPRH-----QPLAFDTEFVNIGDPPDAAGVRCRLPGAYFFSFTLGKL 222

Db 153 KK-----PMSNHYQTVIFDTEFVNLYDHFNNFTGKFCYCVPGLYFFSLNVHTW 202

Qy 223 PRKTLVSKMKRDEVOAMTYDDGASRRREMOSQSVLALRRGDVWVLLSDHDGYGAYS 282

Db 203 NOKETYLHIMKNEEVVILFAQVG--DRSIMOSQSLMELREDOQVWVRLYKGERENAIF 260

Qy 283 NH--GKYTFSGFLY 295

Db 261 SEELDTYITFSGLY 275

RESULT 9

US-09-911-1768-2

Sequence 2, Application US/09911768

Patent No. 6518403

GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.

TITLE OF INVENTION: ANTIBODIES THAT BIND AN

FILE REFERENCE: 97-30D1

CURRENT APPLICATION NUMBER: US/09/911,1768

CURRENT FILING DATE: 2001-07-23

PRIOR APPLICATION NUMBER: 09/118,408

PRIOR FILING DATE: 1998-07-17

PRIOR APPLICATION NUMBER: 60/053,154

PRIOR FILING DATE: 1997-07-18

NUMBER OF SEQ ID NOS: 52  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 281  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-911-176B-2

Query Match 13.1%; Score 215; DB 4; Length 281;

Best Local Similarity 30.2%; Pred. No. 6,6e-14;  
Matches 77; Conservative 26; Mismatches 86; Indels 66; Gaps 11;

53 RCRVGAAYFFSTAGAPKHSLSVLMVRNDEVQALAFDEQRPGARRAASQAMQLDY 112  
75 RCDPOTSMYPATA--VPQINITILKGEKDR-----GDRG-----LQKX 113  
113 GDT--VWLRLHGAAP--HYALGAPGATFSGYLVVADADADAPARGPAPPRPSAAR 167  
114 GKTGSAGARGHTGPKQKSMGAPGERCKSH--YA-----AFSVGR 152  
168 TRSLVGSADAGPGRH-----QPLAFDTEFVNIGDGFDAAGVFRCLPGAYFFSFTLGL 222  
153 KK-----PMHSNHYYQTVIFDTEFVNLYDHFNMFTGKFCYCVPLGYFFSLNVHTW 202  
223 PKRTLSVKLMKRDEVOAMITYDDGASRRRMOSSVMLALRRGDAVWLSSHDDGYGAS 282  
203 NQKETYLMHMKNEEVEVLVFAQVG--DRSIMQSLSIMLELRQDQVWVRLYKGERENALF 260  
283 NH--GKYITFSGFLV 295  
261 SEELDTYITFSGYLV 275

RESULT 10  
US-09-619-740-2

Sequence 2, Application US/09619740  
Patent No. 6544946  
GENERAL INFORMATION:  
APPLICANT: Sheppard, Paul O.  
APPLICANT: Laser, Gerald W.  
APPLICANT: Bishop, Paul D.  
TITLE OF INVENTION: INHIBITORS FOR USE IN HEMOSTASIS AND IMMUNE FUNCTION  
FILE REFERENCE: 99-112C3  
CURRENT APPLICATION NUMBER: US/09/619,740  
CURRENT FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 09/253,604  
PRIOR FILING DATE: 1999-02-19  
PRIOR APPLICATION NUMBER: 09/444,794  
PRIOR FILING DATE: 1999-11-22  
PRIOR APPLICATION NUMBER: 09/506,855  
PRIOR FILING DATE: 2000-02-17  
NUMBER OF SEQ ID NOS: 55  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 281  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-619-740-2

Query Match 13.1%; Score 215; DB 4; Length 281;

Best Local Similarity 30.2%; Pred. No. 6,6e-14;  
Matches 77; Conservative 26; Mismatches 86; Indels 66; Gaps 11;

53 RCRVGAAYFFSTAGAPKHSLSVLMVRNDEVQALAFDEQRPGARRAASQAMQLDY 112  
75 RCDPOTSMYPATA--VPQINITILKGEKDR-----GDRG-----LQKX 113  
113 GDT--VWLRLHGAAP--HYALGAPGATFSGYLVVADADADAPARGPAPPRPSAAR 167  
114 GKTGSAGARGHTGPKQKSMGAPGERCKSH--YA-----AFSVGR 152  
168 TRSLVGSADAGPGRH-----QPLAFDTEFVNIGDGFDAAGVFRCLPGAYFFSFTLGL 222

153 KK-----PMHSNHYYQTVIFDTEFVNLYDHFNMFTGKFCYCVPLGYFFSLNVHTW 202  
223 PKRTLSVKLMKRDEVOAMITYDDGASRRRMOSSVMLALRRGDAVWLSSHDDGYGAS 282  
203 NQKETYLMHMKNEEVEVLVFAQVG--DRSIMQSLSIMLELRQDQVWVRLYKGERENALF 260  
283 NH--GKYITFSGFLV 295  
261 SEELDTYITFSGYLV 275

RESULT 11  
US-09-506-852-2

Sequence 2, Application US/09506852  
Patent No. 6566499  
GENERAL INFORMATION:  
APPLICANT: Sheppard, Paul O.  
APPLICANT: Adipocyte-Specific Protein Homologs  
TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS  
FILE REFERENCE: 97-30  
CURRENT APPLICATION NUMBER: US/09/506,852  
CURRENT FILING DATE: 2000-02-17  
EARLIER APPLICATION NUMBER: 60/053,154  
EARLIER FILING DATE: 1997-07-18  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 281  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-506-852-2

Query Match 13.1%; Score 215; DB 4; Length 281;

Best Local Similarity 30.2%; Pred. No. 6,6e-14;  
Matches 77; Conservative 26; Mismatches 86; Indels 66; Gaps 11;

53 RCRVGAAYFFSTAGAPKHSLSVLMVRNDEVQALAFDEQRPGARRAASQAMQLDY 112  
75 RCDPOTSMYPATA--VPQINITILKGEKDR-----GDRG-----LQKX 113  
113 GDT--VWLRLHGAAP--HYALGAPGATFSGYLVVADADADAPARGPAPPRPSAAR 167  
114 GKTGSAGARGHTGPKQKSMGAPGERCKSH--YA-----AFSVGR 152  
168 TRSLVGSADAGPGRH-----QPLAFDTEFVNIGDGFDAAGVFRCLPGAYFFSFTLGL 222  
153 KK-----PMHSNHYYQTVIFDTEFVNLYDHFNMFTGKFCYCVPLGYFFSLNVHTW 202  
223 PKRTLSVKLMKRDEVOAMITYDDGASRRRMOSSVMLALRRGDAVWLSSHDDGYGAS 282  
203 NQKETYLMHMKNEEVEVLVFAQVG--DRSIMQSLSIMLELRQDQVWVRLYKGERENALF 260  
283 NH--GKYITFSGFLV 295  
261 SEELDTYITFSGYLV 275

RESULT 12  
US-09-336-536-4

Sequence 4, Application US/09336536  
Patent No. 6406884  
GENERAL INFORMATION:  
APPLICANT: Leiby, K.  
APPLICANT: McKay, S.  
APPLICANT: Boscione, S.  
TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF  
FILE REFERENCE: 7853-144  
CURRENT APPLICATION NUMBER: US/09/336,536  
CURRENT FILING DATE: 1999-06-18  
NUMBER OF SEQ ID NOS: 75  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 228  
TYPE: PRT

```
; ORGANISM: Homo sapiens
US-09-336-536-4

Query Match      13.0%; Score 213; DB 4; Length 228;
Best Local Similarity 34.9%; Pred. No. 7.9e-14;
Matches 66; Conservative 18; Mismatches 83; Indels 22; Gaps 6;

QY 122 GAPHYALGAPGATFSG---YLVYADAD---ADAPARGPAPP-----EPRSAGSAATR 169
DB 42 GAP-----GAPGKGGGGRGRLPGPRGDPGPRGEGAPAGTGPAGGCSVPPRSASFASKSE 97
QY 170 SLVGS DAGPRHOPPLAFDTEFVNIGDPPDAAGVFRCLPGAYFFSFTLGKLPKRTLSV 229
DB 98 SRV-----PPPSDAPLPFDRLVNVNQGHDYAVTGKFTQGVGVYFA-VHATVYRASLQF 151
QY 230 KLMKRDVQAMITYDDGASRRREMOSQSVMLALRRGDAVWLISHDHGAYSNHGKYYT 289
DB 152 DLVKNGESIASFPFGGMPKPRASISGAMVRLPEPDQVWGVGDYIGIYASIKTST 211
QY 290 FSGFLVYPD 298
DB 212 FSGFLVYSD 220

RESULT 13
US-09-140-804-2
; Sequence 2, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Humes, Jacqueline M.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT APPLICATION NUMBER: US/09/140,804
; CURRENT FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: 60/056,983
; EARLIER FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-140-804-2

Query Match      13.0%; Score 213; DB 3; Length 243;
Best Local Similarity 34.9%; Pred. No. 8.6e-14;
Matches 66; Conservative 18; Mismatches 83; Indels 22; Gaps 6;

QY 122 GAPHYALGAPGATFSG---YLVYADAD---ADAPARGPAPP-----EPRSAGSAATR 169
DB 57 GAP-----GAPGKGGGGRGRLPGPRGDPGPRGEGAPAGTGPAGGCSVPPRSASFASKSE 112
QY 170 SLVGS DAGPRHOPPLAFDTEFVNIGDPPDAAGVFRCLPGAYFFSFTLGKLPKRTLSV 229
DB 113 SRV-----PPPSDAPLPFDRLVNVNQGHDYAVTGKFTQGVGVYFA-VHATVYRASLQF 166
QY 230 KLMKRDVQAMITYDDGASRRREMOSQSVMLALRRGDAVWLISHDHGAYSNHGKYYT 289
DB 167 DLVKNGESIASFPFGGMPKPRASISGAMVRLPEPDQVWGVGDYIGIYASIKTST 226
QY 290 FSGFLVYPD 298
DB 227 FSGFLVYSD 235

RESULT 14
US-09-336-536-3
; Sequence 3, Application US/09336536
; Patent No. 6406884
; GENERAL INFORMATION:
; APPLICANT: Leiby, K.
; APPLICANT: McKay, C.
```

```
; APPLICANT: Bossone, S.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-144
; CURRENT APPLICATION NUMBER: US/09/336,536
; CURRENT FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-336-536-3

Query Match      13.0%; Score 213; DB 4; Length 243;
Best Local Similarity 34.9%; Pred. No. 8.6e-14;
Matches 66; Conservative 18; Mismatches 83; Indels 22; Gaps 6;

QY 122 GAPHYALGAPGATFSG---YLVYADAD---ADAPARGPAPP-----EPRSAGSAATR 169
DB 57 GAP-----GAPGKGGGGRGRLPGPRGDPGPRGEGAPAGTGPAGGCSVPPRSASFASKSE 112
QY 170 SLVGS DAGPRHOPPLAFDTEFVNIGDPPDAAGVFRCLPGAYFFSFTLGKLPKRTLSV 229
DB 113 SRV-----PPPSDAPLPFDRLVNVNQGHDYAVTGKFTQGVGVYFA-VHATVYRASLQF 166
QY 230 KLMKRDVQAMITYDDGASRRREMOSQSVMLALRRGDAVWLISHDHGAYSNHGKYYT 289
DB 167 DLVKNGESIASFPFGGMPKPRASISGAMVRLPEPDQVWGVGDYIGIYASIKTST 226
QY 290 FSGFLVYPD 298
DB 227 FSGFLVYSD 235

RESULT 15
US-09-686-838B-2
; Sequence 2, Application US/09686838B
; Patent No. 6482612
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Humes, Jacqueline M.
; TITLE OF INVENTION: Adipocyte-Specific Protein Homologs
; FILE REFERENCE: 97-49D1
; CURRENT APPLICATION NUMBER: US/09/686,838B
; CURRENT FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/140,804
; PRIOR FILING DATE: 1998-08-26
; PRIOR APPLICATION NUMBER: US 60/056,983
; PRIOR FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-686-838B-2

Query Match      13.0%; Score 213; DB 4; Length 243;
Best Local Similarity 34.9%; Pred. No. 8.6e-14;
Matches 66; Conservative 18; Mismatches 83; Indels 22; Gaps 6;

QY 122 GAPHYALGAPGATFSG---YLVYADAD---ADAPARGPAPP-----EPRSAGSAATR 169
DB 57 GAP-----GAPGKGGGGRGRLPGPRGDPGPRGEGAPAGTGPAGGCSVPPRSASFASKSE 112
QY 170 SLVGS DAGPRHOPPLAFDTEFVNIGDPPDAAGVFRCLPGAYFFSFTLGKLPKRTLSV 229
DB 113 SRV-----PPPSDAPLPFDRLVNVNQGHDYAVTGKFTQGVGVYFA-VHATVYRASLQF 166
QY 230 KLMKRDVQAMITYDDGASRRREMOSQSVMLALRRGDAVWLISHDHGAYSNHGKYYT 289
DB 167 DLVKNGESIASFPFGGMPKPRASISGAMVRLPEPDQVWGVGDYIGIYASIKTST 226
```

Qy	290	FSGFLVYPD	298
Db	227	FSGFLVYSD	235

Search completed: January 12, 2004, 08:20:04  
Job time : 16.9311 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:09 ; Search time 9.84032 Seconds

(without alignments)  
1661.397 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_160\_358

Perfect score: 894

Sequence: 1 ADADAPARGPPAPPPRPSAF.....LVYDLPAPAPPGIGASELL 170

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.76:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	208.5	23.3	219	2	hypothetical prote
2	182	20.4	253	2	complement subcomp
3	177.5	19.9	215	2	hbernation-relate
4	175	19.6	244	2	gelatin-binding 28
5	170	19.0	245	1	complement subcomp
6	166.5	18.6	674	2	collagen alpha 1(X
7	165.5	18.5	253	2	complement protein
8	165.5	18.5	744	1	collagen alpha 1(X
9	164	18.3	680	1	collagen alpha 1(X
10	164	18.3	680	2	collagen alpha 1(X
11	163.5	18.3	744	1	collagen alpha 1(X
12	163	18.2	224	2	collagen alpha 1(X
13	163	18.2	674	2	cerebellin-like g1
14	160.5	18.0	744	2	collagen alpha 1(X
15	155.5	17.4	246	2	collagen alpha 1(X
16	154.5	17.3	743	2	complement subcomp
17	153	17.1	193	2	collagen alpha 1(X
18	150.5	16.8	215	2	cerebellin precurs
19	144	16.1	253	2	complement C1q B C
20	143	16.0	196	2	hbernation-relate
21	135.5	15.2	245	1	hbernation-relate
22	134	15.0	245	2	complement subcomp
23	132.5	14.8	635	2	collagen alpha 2(V
24	129.5	14.4	170	2	collagen alpha 2(V
25	129	14.4	992	2	hypothetical prote
26	123.5	13.8	423	2	collagen precursor
27	85	9.5	589	2	conserved hypotet
28	85	9.5	1228	2	multimerin, endoch
29	82.5	9.2	472	2	hypothetical prote

30	82.5	9.2	1433	2	A46053	bullous pemphigoid
31	81.5	9.1	420	2	E75130	hypothetical prote
32	81	9.1	618	2	T24228	hypothetical prote
33	80.5	9.0	216	2	T29039	hypothetical prote
34	80	8.9	657	2	E70025	probable acylamino
35	78.5	8.8	602	2	T35782	probable secreted
36	76.5	8.6	425	2	T41172	hypothetical prote
37	76.5	8.6	706	2	S33761	transferrin precur
38	76	8.5	2796	2	UC4743	fatty-acid synthas
39	75.5	8.4	175	2	B83411	hypothetical prote
40	75.5	8.4	250	2	S27918	coat protein - gar
41	75.5	8.4	250	2	J02181	viral coat 27.4K p
42	75.5	8.4	400	2	B72650	hypothetical prote
43	75.5	8.4	493	2	E71725	glutamy1-tRNA amid
44	75.5	8.4	660	2	T03038	probable inhibitor
45	75.5	8.4	1445	2	A59437	KIAA1204 protein 1

#### ALIGNMENTS

##### RESULT 1

hypothetical protein DKFZp586B0621.1 - human (fragment)

C/Species: Homo sapiens (man)

C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 02-Sep-2000

C/Accession: T14782

R/Ottensmeyer, B.; Obermayer, B.; Mewes, H.W.; Gassenhuber, J.; Wilmann, S.

submitted to the Protein Sequence Database, August 1999

A/Reference number: Z18184

A/Accession: T14782

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-219 <OTT>

A/Cross-references: EMBL:AL110261

A/Experimental source: adult uterus; clone DKFZp586B0621

C/Genetics:

A/Note: DKFZp586B0621.1

C/Superfamily: complement C1q carboxyl-terminal homology

Query Match 23.3%; Score 208.5; DB 2; Length 219;

Best Local Similarity 35.7%; Pred. No. 1,4e-12;

Matches 56; Conservative 18; Mismatches 72; Indels 11; Gaps 3;

QY	4	DAPARGPAPP-----EPSSAFSAARTSRVSGDAGPGRHOPIAFPTFEVNIIGDPDA	58
DB	61	EAGPAGPTGAGGECVPPRASFSAKRSERV-----PPSDAPLPFDRLVNEQHYDAV	115
QY	59	AGVFRCLPGAYFPFSTLGLPKRTLSVKLMKRDVQAMVDDGASRRRMOQSVMVA	118
DB	116	TGKFTGVPGVYVFA-VHATVVRASLQFDLVKNGESIASFPFGGMPKFAISGGMVR	174
QY	119	LRRGDVWLISHDDGAYSNHGKTYTTFSGFLVYD	155
DB	175	LEPEDVWVGVDYTGIVASIKTSTSGFLVYD	211

##### RESULT 2

complement subcomponent C1q chain B precursor [validated] - human

N/Alternate names: complement subcomponent C1q beta chain

C/Species: Homo sapiens (man)

C/Date: 22-May-1981 #sequence\_revision 31-May-1996 #text\_change 08-Dec-2000

A/Accession: B23422; A23422; B90304; A90301; B90315; A03206

R/Reid, K.B.M.

Biochem. J. 231, 729-735, 1985

A/Title: Molecular cloning and characterization of the complementary DNA and gene codin

A/Reference number: A23422; MUID:86076906; PMID:3000358

A/Accession: B23422

A/Molecule type: DNA

A/Residues: 'HS', 1-32 <RET>

A/Note: the authors translated the codon AGT for the second position as Arg; they were

A:Molecule type: mRNA  
 A:Residues: 28-253 <RE1>  
 A:Cross-references: EMBL:X03084  
 A>Note: the authors translated the codon ACA for residue 46 as Ile  
 R.Reid, K.B.M.  
 Biochem. J. 179, 367-371, 1979  
 A>Title: Complete amino acid sequences of the three collagen-like regions present in sub  
 A:Reference number: A90304; MUID:80020137; PMID:486087  
 A:Accession: B90304  
 A:Molecule type: protein  
 A:Residues: 'E', 29-84, 'D', 86-99, 'P', 101-135 <RE5>  
 R.Reid, K.B.M.; Thompson, E.O.P.  
 Biochem. J. 173, 863-868, 1978  
 A>Title: Amino acid sequence of the N-terminal 108 amino acid residues of the B chain of  
 A:Reference number: A90301; MUID:79041552; PMID:708376  
 A:Accession: A90301  
 A:Molecule type: protein  
 A:Residues: 28-99, 'P', 101-195 <RE3>  
 R.Reid, K.B.M.; Gagnon, J.; Frampton, J.  
 Biochem. J. 203, 559-569, 1982  
 A>Title: Completion of the amino acid sequences of the A and B chains of subcomponent C1  
 A:Reference number: A90315; MUID:82283890; PMID:6981411  
 A:Accession: B90315  
 A:Molecule type: protein  
 A:Residues: 136-253 <RE4>  
 A>Note: 176-Glx may also be present  
 C:Comment: The first component of complement is a calcium-dependent complex of the three  
 A:Reference number: C1q (enzyme), C1s (proenzyme), and the other eight components of complement.  
 C:Comment: The C1q subcomponent is composed of nine subunits, six of which are disulfide  
 (see PIR:CIHQCC) chain. Equimolar amounts of the A, B, and C chains are found after red  
 A:Gene: GDB:C10B  
 A:Cross-references: GDB:119043; OMIM:120570  
 A:Map position: 1p36.3-1p34.1  
 C:Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal hom  
 C:Keywords: complement pathway; glycoprotein; heterodimer; hydroxylysine; hydroxyproline  
 F:1-27/Domain: signal sequence #status predicted <SIG>  
 F:28-253/Product: complement subcomponent C1q chain B #status experimental <MAT>  
 F:123-116/Domain: collagenous, triple helix <COL>  
 F:123-249/Domain: complement C1q carboxyl-terminal homology <CIQ>  
 F:128/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experim  
 F:31/Disulfide bonds: interchain (to chain A-26) #status experimental  
 F:33,38,41,55,65,83,86,101,104,107/Modified site: 4-hydroxyproline (Pro) #status expe  
 F:59,62,77,92,99,110/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F:59,62,98,110/Binding site: carbohydrate (Lys) (covalent) #status experimental

Query Match 20.4%; Score 182; DB 1; Length 253;  
 Best Local Similarity 29.1%; Pred. No. 5, 8e-10;  
 Matches 48; Conservative 32; Mismatches 61; Indels 24; Gaps 5;

QY 6 PARGPAPPEP-----RSFSAARTSLVSGDAGPRHOPLAFTPEFVNG 52  
 DB 97 PFGGPGAPGAPGPKGSGGYKATQKAFSATRTIN-----PLRDDQTRRDHYITNN 150  
 QY 53 GPDFAAAGVRCRLPGAVFESFTLGLPKRTLSVKLMKNRDEVQAMI-YDQASRRRMO 111  
 DB 151 NNVEPSGKFTCKVPELVYFTTHASS--RGNLQVNLMRERAKQKVTCDVAVMTFCYT 208  
 QY 112 SOSVWLALRRGDAVWLSHDHDGAYVSNHGKTYTFSGLVYVDL 156  
 DB 209 TCGMVVLKLEGGENVFLQATDKN--SLTMEGANSIFSGLLPDM 251

RESULT 3  
 B48150  
 hibernation-related protein HP-25 precursor - Siberian chipmunk  
 C:Species: Eutamias sibiricus (Siberian chipmunk)  
 C>Date: 16-Feb-1994 #sequence\_revision 25-Apr-1997 #ext\_change 20-Jun-2000  
 C:Accession: B48150; B41752  
 R.Jakamatsu, N.; Ohba, K.; Kondo, J.; Kondo, N.; Shiba, T.  
 Mol. Cell. Biol. 13, 1516-1521, 1993  
 A>Title: Hibernation-associated gene regulation of plasma proteins with a collagen-like  
 A:Reference number: A48150; MUID:93180798; PMID:8441393

A:Accession: B48150  
 A:Molecule type: protein  
 A:Residues: 1-215 <TRK>  
 A:Cross-references: GB:D12975; NID:9287469; PIDN:BA02352.1; PID:9287470  
 A>Note: the source is designated as Tamias asiaticus in Genbank entry TMSHP25, release  
 A>Note: sequence extracted from NCBI backbone (NCBIN:125946, NCBIPI:125947)  
 J. Kondo, N.; Kondo, J.  
 J. Biol. Chem. 267, 473-478, 1992  
 A>Title: Identification of novel blood proteins specific for mammalian hibernation.  
 A:Reference number: A41752; MUID:92112696; PMID:1730610  
 A:Accession: B41752  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 29-62;84-130;172-183;187-192;201-215 <KON>  
 C:Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal ho  
 C:Keywords: glycoprotein; hibernation; plasma  
 F:1-28/Domain: signal sequence #status predicted <SIG>  
 F:29-215/Product: hibernation-related protein HP-25 #status experimental <MAT>  
 F:91-214/Domain: complement C1q carboxyl-terminal homology <CIQ>  
 F:167/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 19.8%; Score 177.5; DB 2; Length 215;  
 Best Local Similarity 32.4%; Pred. No. 1.3e-09;  
 Matches 58; Conservative 21; Mismatches 49; Indels 51; Gaps 10;

QY 5 AP-ARGPPAPPEP-----EPRSAFSAARTSLVSGDAGPRHOPLA 43  
 DB 56 APGLAGPPGPGVCGIPGPGPDVCKSGSRPFAVAVKL-----SERPPEP-FQPIV 108  
 QY 44 FDTFVNIGDFDPAAGVRCRLPGAVFESFTLGLPKRTLSVKLMKNRDEVQAMITDDG 103  
 DB 109 FKEALYNQEGHFNMATGEFSCVLPGVNFGFDI-RLFQSSVKRLM--RQGIQV----- 159  
 QY 104 ASRRREMOSQ-----SVWLALRRGDAVWLSHDHDGAYVSNHG-KYTFSGFLVY 153  
 DB 160 --REKEADNDSYHAMGSYVLMAGKQDKVLSKLR---CTSEKGIITHVVFQYLYL 213

RESULT 4  
 J04708  
 gelatin-binding 28k protein precursor - human  
 N:Alternate names: adipose specific collagen-like factor  
 C:Species: Homo sapiens (man)  
 C>Date: 10-May-1996 #sequence\_revision 19-Jul-1996 #ext\_change 20-Sep-1999  
 C:Accession: J04708; J04944  
 R.Maeda, K.; Okubo, K.; Shimomura, I.; Funahashi, T.; Matsuzawa, Y.; Matsubara, K.  
 Biochem. Biophys. Res. Commun. 221, 286-289, 1996  
 A>Title: cDNA cloning and expression of a novel adipose specific collagen-like factor.  
 A:Reference number: J04708; MUID:96224171; PMID:8619847  
 A:Accession: J04708  
 A:Molecule type: mRNA  
 A:Residues: 1-244 <MAE>  
 A:Cross-references: DDBJ:D45371; NID:9871886; PIDN:BA08227.1; PID:9871887  
 A:Experimental source: adipose tissue  
 R.Nakano, Y.; Tohe, T.; Choi-Miyura, N.H.; Maeda, T.; Tomita, M.  
 J. Biochem. 120, 803-812, 1996  
 A>Title: Isolation and characterization of GBP28, a novel gelatin-binding protein purif  
 A:Reference number: J04944; MUID:97103474; PMID:8947845  
 A:Accession: J04944  
 A:Molecule type: protein  
 A:Residues: 19-38;93-100;101-112;135-149;173-178 <NAK>  
 C:Comment: This protein is an endogenous factor that binds with a collagen-like domain  
 C:Genetics:  
 A:Gene: apm1  
 C:Superfamily: unassigned collagens; complement C1q carboxyl-terminal homology  
 C:Keywords: adipose tissue; glycoprotein; hydroxyproline  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-244/Product: gelatin-binding 28kDa protein #status experimental <MAT>  
 F:42-107/Region: collagen-like  
 F:114-241/Domain: complement C1q carboxyl-terminal homology <CIQ>  
 F:95/Modified site: 4-hydroxyproline (Pro) #status experimental  
 F:230/Binding site: carbohydrate (Asn) (covalent) #status predicted



Query Match 19.6%; Score 175; DB 2; Length 244;  
 Best Local Similarity 31.1%; Pred. No. 2.6e-09;  
 Matches 51; Conservative 24; Mismatches 65; Indels 24; Gaps 6;

QY 7 ARGPAPP-----EP-----RSAPSAARTSLVSGDAGPGRHQPPLAFDTEFVNIG 52  
 |||||  
 DB 88 AEGPRGPGFQGRKGEGBEAGVYRSASF-----VGLERYVTIPMPFRFTIFINQ 140  
 |||||

QY 53 GDDPAAAGVRCRLPGAYFFSFTLGKLPRTLSVKMKNDVQAMTYDDGASRRMQS 112  
 |||||  
 DB 141 NHVDGSGKFRHCNIPGLYFAYNI-TYMKDVKSLEK-KDKAMLFYDQYQENNVQAS 198  
 |||||

QY 113 QSWALATRGDAVWLISH-DHDDGAGSNNKITYTSGFLVYPD 155  
 |||||  
 DB 199 GSVLLHLEVGQVWLQYVGEGRNGLYADNDNSTFGFLYHD 242  
 |||||

RESULT 5  
 C1H1QC  
 Complement subcomponent C1q chain C precursor - human  
 N/Alternate names: Complement subcomponent C1q gamma chain  
 C/Species: Homo sapiens (man)  
 C/Date: 22-May-1981 #sequence\_revision 31-May-1996 #text\_change 22-May-1998  
 C/Accession: S14351; A03207  
 R/Seller: G.C.; Blake, D.J.; Reid, K.B.M.  
 Biochem. J. 274, 481-490, 1991  
 A/Title: Characterization and organization of the genes encoding the A-, B- and C-chains  
 A/Reference number: S14350; MUID:91174759; PMID:11706597  
 A/Accession: S14351  
 A/Status: not compared with conceptual translation  
 A/Molecule type: DNA  
 A/Residues: 1-245 <SEL>  
 R/Reid, K.B.M.  
 Biochem. J. 179, 367-371, 1979  
 A/Title: Complete amino acid sequences of the three collagen-like regions present in sub  
 A/Reference number: A90304; MUID:80020137; PMID:486087  
 A/Accession: A03207  
 A/Molecule type: protein  
 A/Residues: 29-56, 'P', 58-65, 'K', 67-71, 'P', 73-83, 'K', 85-86, 'D', 88-89, 'N', 91-122 <REI>  
 C/Comment: The first component of complement is a calcium-dependent complex of the three  
 C/Keyword: complement, C1r (enzyme), C1s (proenzyme), and the other eight components of complement.  
 C/Comment: The C1q subcomponent is composed of nine subunits, six of which are disulfide  
 C/Comment: dimers of the C chain. Equimolar amounts of the A, B, and C chains are found after red  
 C/Genetics:  
 A/Gene: GDB: 128132; OMIM:120575  
 A/Cross-references: GDB:128132; OMIM:120575  
 A/Map position: 1p36.3-1p34.1  
 A/Intons: 60/3  
 C/Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal hom  
 C/Keyword: complement subcomponent C1q chain A; complement C1q carboxyl-terminal hom  
 C/Keyword: complement pathway; glycoprotein; homodimer; hydroxyllysine; hydroxyproline;  
 F/1-28/Domain: signal sequence #status predicted <SIG>  
 F/29-245/Product: complement subcomponent C1q chain B #status predicted <MAT>  
 F/31-114/Domain: collagens, triple helix <COL>  
 F/121-244/Domain: complement C1q carboxyl-terminal homology <C1O>  
 F/32/Disulfide bonds: Interchain #status experimental  
 F/36,39,42,45,54,63,81,93,96,105/Modified site: 4-hydroxyproline (Pro) #status experi  
 F/57,72,75/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F/75/Binding site: carbohydrate (Lys) (covalent) #status experimental

Query Match 19.0%; Score 170; DB 1; Length 245;  
 Best Local Similarity 26.6%; Pred. No. 7.9e-09;  
 Matches 46; Conservative 27; Mismatches 56; Indels 44; Gaps 6;

QY 9 GPPA-----PEP-----RSAPSAARTSLVSGDAGPGRHQP-----L 42  
 |||||  
 DB 91 GPPGMGVPPMGIPGEPSEGRYKQKFGVFTVTR-----QTHQPPANSLI 138  
 |||||

QY 43 APTEFVNIGDPAAGVRCRLPGAYFFSFTLGKLPRTLSVKMKNDVQAMTYDD 102  
 |||||  
 DB 139 RFNAVLTNPGGDYDTSGKFTCKVPGLYFVYHAS---HTANLCVLTALRSRGKVVTFPG 194  
 |||||

QY 103 GASRRREMOSQSWMLALRGDAVWLISHDHDDGAGSNNKITYTSGFLVYPD 155  
 |||||

DB 195 HTSKTNQVNSGVTLRLQVGEVWLA VNDY--YDNVIGQSDSVFSGFLFPD 245

RESULT 6  
 S23297  
 Collagen alpha 1(X) chain precursor - chicken  
 N/Alternate names: type X collagen  
 C/Species: Gallus gallus (chicken)  
 C/Date: 07-Oct-1994 #sequence\_revision 10-Nov-1995 #text\_change 13-Aug-1999  
 C/Accession: S23297; A31896; S65594; S77711; I50218  
 R/Niomiya, Y.; Castagnola, P.; Gerecke, D.; Gordon, M.K.; Jacenko, O.; LuValle, P.; Mc  
 maghugh, N.; Olsen, B.R.  
 In Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp.79-114, Academic Pr  
 A/Title: The molecular biology of collagens with short triple-helical domains.  
 A/Reference number: S22243  
 A/Accession: S23297  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-674 <NIN>  
 R/LuValle, P.; Niomiya, Y.; Rosenblum, N.D.; Olsen, B.R.  
 J. Biol. Chem. 263, 18378-18385, 1988  
 A/Title: The type X collagen gene. Intron sequences split the 5'-untranslated region an  
 A/Reference number: A31896; MUID:89054019; PMID:2461368  
 A/Accession: A31896  
 A/Molecule type: mRNA  
 A/Residues: 1-75 <LUV>  
 R/Niomiya, Y.; Gordon, M.; van der Rest, M.; Schmid, T.; Linsemayer, T.; Olsen, B.R.  
 J. Biol. Chem. 261, 5041-5050, 1986  
 A/Title: The developmentally regulated type X collagen gene contains a long open reading  
 A/Reference number: I50218; MUID:86168227; PMID:3082876  
 A/Accession: S65594  
 A/Molecule type: DNA  
 A/Residues: 'T', '9', 'D', '11-12', 'EDQKLYLFTM', '30-31', 'TCKSGRAFTTMIQNMADLVASH', '48-89', 'L',  
 629', 'PQAVLSLSMTRIKGSSCOIQNPVSPILMFILISQVLSLNNPLTMS' <NINI>  
 A/Cross-references: EMBL:M13496; NID:g211659; PIDN:AAA48736.1; PID:g211700  
 A/Accession: S77711  
 A/Molecule type: protein  
 A/Residues: 104-112, 'X', 114-117, 453-466 <NIN2>  
 C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
 C/Keyword: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
 C/Keyword: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology; hydroxyproline;  
 F/1-16/Domain: signal sequence #status predicted <SIG>  
 F/47-673/Domain: complement C1q carboxyl-terminal homology <C1O>  
 F/453/456/Modified site: hydroxyproline (Pro) #status experimental  
 F/611/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 18.6%; Score 166.5; DB 2; Length 674;  
 Best Local Similarity 28.9%; Pred. No. 5.8e-08;  
 Matches 50; Conservative 19; Mismatches 71; Indels 33; Gaps 5;

QY 9 GPPAPP-----RSAPSAARTSLVSGDAGPGR 38  
 |||||  
 DB 501 GPPGPPGPGGSGTIRGKYKGRSLSGMSFMKAQANQALTMPSAFVILISKAYPGAT 560  
 |||||

QY 39 HOPPLAFTEFVNIGDPAAGVRCRLPGAYFFSFTLGKLPRTLSVKMKNDVQAM 98  
 |||||  
 DB 561 -VPITFDKLTLYNRQOHYPRTGIFTCRIPLGLYFVYH-HAKGTVMVWALYKNGSPVM-Y 617  
 |||||

QY 99 IYDDASRRREMOSQSWMLALRGDAVWLISHDHDDGAGSNNKITYTSGFL 151  
 |||||  
 DB 618 TYDEYQKGYLDASGSAVIDLMENDQVWLQLPNSNGILYSEYVHSSFGFL 670  
 |||||

RESULT 7  
 S49158  
 Complement protein C1q beta chain precursor - rat  
 C/Species: Rattus norvegicus (Norway rat)  
 C/Date: 16-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 20-Aug-1999  
 C/Accession: S49158  
 R/Schwaible, W.; Petry, F.; Loos, M.  
 Submitted to the EMBL Data Library, March 1993  
 A/Description: cDNA cloning and expression of the mRNA encoding the B-chain of rat C1q.  
 A/Reference number: S49158  
 A/Accession: S49158

A>Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-253 <SCH>  
A:Cross-references: EMBL:X71127, NID:G510191, PIDN:CAA50440.1, PID:G510192  
C:Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal hom  
F:121-249/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 18.5%; Score 165.5; DB 2; Length 253;  
Best Local Similarity 27.6%; Pred. No. 2.2e-08;  
Matches 47; Conservative 32; Mismatches 56; Indels 35; Gaps 7;

8 RGPAPPEPRSS-----AFSAATRSVSGSDAGPRRHOPAFDTEFVNI 53  
96 KGAGPGRGPRKSGSDYKATQKAFSAART---VNSALRP---NQAIRKXITVND 149  
54 DFDAAAGVFCRLPGAFVFSFTLCKLPKTLSTVKLMKRRD---VQAMTYDDGSRREM 110  
150 NYEPRSGKFTCKVGLYFTVYHASS--RGNLCVNIIVGRDRMQKVLTFCDYAQNTFOV 207

111 QSQSVMLARKGDVWMLSHHDGV---GAYSHKKTITFSGFLVYPDL 156  
208 TTGCVLKLKEGEVAVHQLATDKNSLLGVEGANS-----ITFGFLPDM 251

RESULT 8  
S23298  
collagen alpha 1(VIII) chain - chicken  
C/Species: Gallus gallus (chicken)  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C/Accession: S23298  
R:Nomura, Y.; Castagnola, P.; Gerecke, D.; Gordon, M.K.; Jacenko, O.; LuValle, P.; McC  
maguchi, N.; Olsen, B.R.  
In Extracellular Matrix Gene, Sandell L.J. and Boyd C.D., eds., pp 79-114, Academic Pre  
A>Title: The molecular biology of collagens with short triple-helical domains.  
A/Reference number: S22243  
A/Accession: S23298  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-744 <NIN>  
C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
F:617-743/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 18.5%; Score 165.5; DB 1; Length 744;  
Best Local Similarity 25.0%; Pred. No. 8.2e-08;  
Matches 47; Conservative 27; Mismatches 65; Indels 49; Gaps 5;

9 GPPAPPEPRSS-----AFSAATRSVSGDAGPG----- 36  
563 GPPGPPGPPAVWPPTAPQGEVLPDMGLGIDGVKTPHVAARK-----GKNGSPAYEMPA 617  
37 -----PRHQPLAPTEFVNIGDPDAAAGVFCRLPGAFVFSFTLCKLPKTLSTV 86  
618 PTAELTAPFPVPGADIKPRLLYNORONYPQTGFTCEVPVYFAHV-HCKGQNVWV 676  
87 KLMKRDEQAMVITDDGASRRREMOSQVMLALRGDAVWMLSHHDGAYSHNGKTYT 146  
677 ALFKNNRPVW-YTYDEYKKGFLDQASGSAVLLRFEDRVFLQNPSEQAAGLYAGQYHSS 735

147 FSGFLVYP 154  
736 FSGVLLYP 743

RESULT 9  
CGHUID  
collagen alpha 1(X) chain precursor - human  
N:Alternate names: procollagen alpha 1(X) chain  
C/Species: Homo sapiens (man)  
C/Date: 22-Nov-1993 #sequence\_revision 03-Nov-1995 #text\_change 22-Jun-1999  
C/Accession: S26396; S30086; S15826; S18249; A43901; I51870; S21856  
R:Reichenberger, E.; Beier, F.; LuValle, P.; Olsen, B.R.; von der Mark, K.; Bestling, W.  
FEBS Lett. 311, 305-310, 1992  
A>Title: Genomic organization and full-length cDNA sequence of human collagen X.

A/Reference number: S26396; MUID:93012005; PMID:1397333  
A/Accession: S26396  
A:Molecule type: DNA  
A:Residues: 1-680 <REI>  
A:Cross-references: EMBL:X68952; EMBL:X72578; EMBL:X72579; EMBL:X72580; GB:S47714; GB:S4  
R/Apte, S.S.  
Submitted to the EMBL Data Library, March 1992  
A/Reference number: S30085  
A/Accession: S30086  
A:Molecule type: DNA  
A:Residues: TTPYGVWCVCL/52-680 <APT>  
A:Cross-references: EMBL:X65120; NID:G23129  
A/Note: the initial difference is probably due to translation of an intronic sequence  
R/Apte, S.; Mattei, M.G.; Olsen, B.R.  
FEBS Lett. 282, 393-396, 1991  
A>Title: Cloning of human alpha-1(X) collagen DNA and localization of the COL10A1 gene  
A/Reference number: S15826; MUID:91243858; PMID:2037056  
A/Accession: S15826  
A:Molecule type: DNA  
A:Residues: 561-647, 'G', 649-666 <AP2>  
A:Cross-references: EMBL:X58879; NID:G30013; PIDN:CAA41686.1; PID:G30014  
R:Thomas, J.T.; Cresswell, C.J.; Rash, B.; Nicolai, H.; Jones, T.; Solomon, E.; Grant, J  
Biochem. J. 280, 617-623, 1991  
A>Title: The human collagen X gene. Complete primary translated sequence and chromosomal  
A/Reference number: S18249; MUID:92109659; PMID:1764025  
A/Accession: S18249  
A:Molecule type: DNA  
A:Residues: 1-26, 'T', 28-680 <THO>  
A:Cross-references: EMBL:X60382; NID:G30094; PIDN:CAA42933.1; PID:G30095  
A/Note: the sequence from Fig. 3 is inconsistent with that from Fig. 2 in having 532-A1  
R:Reichenberger, E.; Aigner, T.; von der Mark, K.; Stoss, H.; Bestling, W.  
Dev. Biol. 149, 562-572, 1991  
A>Title: In situ hybridization studies on the expression of type X collagen in fetal hu  
A/Reference number: A43901; MUID:92077285; PMID:1743401  
A/Accession: A43901  
A:Molecule type: mRNA  
A:Residues: 547-656 <RE2>  
A:Cross-references: GB:M74050; GB:D57494; NID:G339884; PIDN:AA61221.1; PID:G553796  
A/Note: sequence extracted from NCBI backbone (NCBI:G59012, NCBI:G59014)  
R:Wallis, G.A.; Rash, B.; Sweetman, W.A.; Thomas, J.T.; Super, M.; Evans, G.; Grant, M.  
Am. J. Hum. Genet. 55, 169-178, 1994  
A>Title: Amino acid substitutions of conserved residues in the carboxyl-terminal domain  
pe Schmid.  
A/Reference number: I51870; MUID:94136476; PMID:8304336  
A/Accession: I51870  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 520-597, 'D', 599-680 <WAL>  
A:Cross-references: GB:S68531; NID:G545180; PIDN:AA60615.1; PID:G545181  
A/Note: mutant sequence from patient with metaphyseal chondrodysplasia type Schmid  
C/Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
ed and subsequently O-glycosylated.  
C/Genetics:  
A:Gene: GDB:COL10A1  
A:Cross-references: GDB:128635; OMIM:120110  
A:Map position: 6q21-6q22  
A:Introns: 52/1  
A/Note: a defect in this gene may cause Schmid metaphyseal chondrodysplasia  
C/Complex: type X collagen may be a homotrimer  
C/Function:  
A/Description: structural component of extracellular fibrous polymer specifically and t  
be important for skeletogenesis  
C/Supersfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
C/Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer; hydroxyllysine,  
F:1-680/Domain: signal sequence #status predicted <SIG>  
F:19-680/Product: collagen alpha 1(X) chain #status predicted <MAT>  
F:19-56/Domain: amino-terminal nonhelical #status predicted <NC2>  
F:57-519/Region: interrupted helical  
F:520-680/Domain: amino-terminal nonhelical #status predicted <NC1>  
F:553-679/Domain: complement C1q carboxyl-terminal homology <C1Q>  
F:617/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 18.3%; Score 164; DB 1; Length 680;  
 Best Local Similarity 25.6%; Pred. No. 1e-07;  
 Matches 50; Conservative 20; Mismatches 71; Indels 54; Gaps 4;

QY 9 GPPAPPEPRSAFSAARTSLVSGDAGPGRHQ----- 40  
 DB 490 GPPGPPGPR---GHSGBGLPGPPGPPGQAVMPDGF1KAGQRLSGMPLVSANHCV 546  
 QY 41 -----PLAFDTFVNIGDFFDAAGVFCRLPGAYFSSFTLGL 79  
 DB 547 TGMPSAFTVILSKAYPAIGTPIPDKILYRQOHYDPRGTGFTCQIPGIYFSYHV-HV 605  
 QY 80 PRKTLISYKMKNRDEVQAMITYDDGASRRRMOQSVMALRRGDAVWLSHDHGYGAYS 139  
 DB 606 KGTWVWGLYKNGTPTW-YTYDEYKGYLDQAGSALITDENDQVWLQPLNASENGLYS 664  
 QY 140 NHGKYITFSGFLVYP 154  
 DB 665 SEYVHSSFGFLVAP 679

## RESULT 10

collagen alpha 1(X) chain precursor - mouse  
 C/Species: Mus musculus (house mouse)  
 C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 13-Aug-1999  
 C/Accession: S31216, S28807, S22215, S30127, I48299, S26397, S31830  
 R/Kong, R.Y.C.; Kwan, K.M.; Lau, E.T.; Thomas, J.T.; Boot-Handford, R.P.; Grant, M.E.; C  
 Eur. J. Biochem. 213, 99-111, 1993  
 A/Title: Intronic-exon structure, alternative use of promoter and expression of the mouse  
 A/Reference number: S31216; MUID:93238750; PMID:8477738  
 A/Accession: S31216  
 A/Molecule type: DNA  
 A/Residues: 1-680 <KON>  
 A/Cross-references: EMBL:Z21610; NID:949793; PIDN:CAA79736.1; PID:949794  
 R/Elima, K.; Berola, I.; Rosati, R.; Metserant, M.; Garofalo, S.; Peraelae, M.; de Crc  
 Biochem. J. 289, 247-253, 1993  
 A/Title: The mouse collagen X gene: complete nucleotide sequence, exon structure and exp  
 A/Reference number: S28807; MUID:93143676; PMID:8424763  
 A/Accession: S28807  
 A/Molecule type: DNA  
 A/Residues: 1-285, 'A', 287-680 <ELI>  
 A/Cross-references: EMBL:X67348; NID:950480; PIDN:CAA47763.1; PID:950481  
 R/Elima, K.; Metserant, M.; Kallio, J.; Peraelae, M.; Berola, I.; Garofalo, S.; de Crc  
 Biochim. Biophys. Acta 1130, 78-80, 1992  
 A/Title: Specific hybridization probes for mouse alpha-2(I)X and alpha-1(X) collagen mRN  
 A/Reference number: S22215; MUID:92182017; PMID:1543751  
 A/Accession: S22215  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 385-450, 'K', 452-627 <ELA>  
 A/Cross-references: EMBL:X63013; NID:949795; PIDN:CAA44741.1; PID:949796  
 R/Apte, S.S.; Olsen, B.R.  
 Matrix 13, 165-179, 1993  
 A/Title: Characterization of the mouse type X collagen gene.  
 A/Reference number: S30127; MUID:93261348; PMID:8492743  
 A/Accession: S30127  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-112, 'F', 14-26, 'S', 28-247, 'L', 249-285, 'A', 287-305, 'F', 307-416, 'S', 418-499, 'L  
 R/Apte, S.S.; Seldin, M.F.; Hayashi, M.; Olsen, B.R.  
 Eur. J. Biochem. 206, 217-224, 1992  
 A/Title: Cloning of the human and mouse type X collagen genes and mapping of the mouse t  
 A/Reference number: I48299; MUID:92267014; PMID:1587271  
 A/Accession: I48299  
 A/Status: preliminary; translated from GB/EMBL/DBD  
 A/Molecule type: DNA  
 A/Residues: 52-247, 'L', 249-285, 'A', 287-305, 'F', 307-416, 'S', 418-499, 'L', 501-566, 'C', 568, '  
 A/Cross-references: EMBL:X65121; NID:950482; PIDN:CAA46237.1; PID:9567031  
 R/Summers, T.A.; Irwin, M.H.; Mayne, R.; Ballian, G.  
 J. Biol. Chem. 263, 581-587, 1988  
 A/Title: Monoclonal antibodies to type X collagen. Bico-synthetic studies using an antibod  
 A/Reference number: S26397; MUID:88087150; PMID:1826450

A/Accession: S26397  
 A/Molecule type: Protein  
 A/Residues: 'SDGYFSQ', 24-26, 'KQ' <SUM>  
 C/Genetics:  
 A/Gene: Collagen-1  
 A/Map position: 10  
 A/Introns: 51/3  
 C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
 C/Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-680/Product: collagen alpha 1(X) chain #status predicted <MNT>  
 F:553-679/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 18.3%; Score 164; DB 2; Length 680;  
 Best Local Similarity 25.1%; Pred. No. 1e-07;  
 Matches 49; Conservative 22; Mismatches 70; Indels 54; Gaps 4;

QY 9 GPPAPPEPRSAFSAARTSLVSGDAGPGRHQ----- 40  
 DB 490 GPPGPPGPR---GHSGBGLPGPPGPPGQAVMPDGF1KAGQRLSGMPLVSANHCV 546  
 QY 41 -----PLAFDTFVNIGDFFDAAGVFCRLPGAYFSSFTLGL 79  
 DB 547 TGMPSAFTVILSKAYPAVCAPIPFDEILYRQOHYDPRGIFTCKIPGIYFSYHV-HV 605  
 QY 80 PRKTLISYKMKNRDEVQAMITYDDGASRRRMOQSVMALRRGDAVWLSHDHGYGAYS 139  
 DB 606 KGTWVWGLYKNGTPTW-YTYDEYKGYLDQAGSALITDENDQVWLQPLNASENGLYS 664  
 QY 140 NHGKYITFSGFLVYP 154  
 DB 665 SEYVHSSFGFLVAP 679

## RESULT 11

collagen alpha 1(VIII) chain precursor - rabbit  
 C/Species: Oryctolagus cuniculus (domestic rabbit)  
 C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C/Accession: A34246  
 R/Yamaguchi, N.; Banya, P.D.; van der Rest, M.; Ninomiya, Y.  
 J. Biol. Chem. 264, 16022-16029, 1989  
 A/Title: The cloning and sequencing of alpha1(VIII) collagen cDNAs demonstrate that typ  
 omains similar to those of type X collagen.  
 A/Reference number: A34246; MUID:89380199; PMID:2476437  
 A/Accession: A34246  
 A/Molecule type: mRNA  
 A/Residues: 1-744 <YAM>  
 A/Cross-references: GB:005042; NID:9164895; PIDN:AAA31204.1; PID:9164896  
 C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
 F:1-20/Domain: signal sequence #status predicted <SIG>  
 F:21-744/Product: collagen alpha 1(VIII) chain #status predicted <MNT>  
 F:21-117/Region: amino-terminal nonhelical  
 F:118-571/Region: interrupted helical  
 F:572-744/Region: carboxyl-terminal nonhelical  
 F:617-743/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 18.3%; Score 163.5; DB 1; Length 744;  
 Best Local Similarity 25.0%; Pred. No. 1.3e-07;  
 Matches 47; Conservative 26; Mismatches 66; Indels 49; Gaps 5;

QY 9 GPPAPPEPRSAFSAARTSLVSGDAGPGRHQ----- 36  
 DB 563 GPPGPPGPRVMPPTAPGGEYLPDGLGDKYKTHAIAAKR-----GKNGSPAYEMPA 617  
 QY 37 -----PRHPIAFDTFVNIGDFFDAAGVFCRLPGAYFSSFTLGL 86  
 DB 618 FTAELTAPPPEPVGAPFKFRLLYNGRONYNPCGIFTCEVPYYPAYHV-HCKGQVWV 676  
 QY 87 KLMKRNDEVQAMITYDDGASRRRMOQSVMALRRGDAVWLSHDHGYGAYSNHGKYIT 146  
 DB 677 ALFKNNEPWA-YTYDEYKGYLDQAGSALLRLPDRVFLQMPSEQAAGLYAGQYVHSS 735

QY 147 FSGFLVYP 154  
|||:|  
Db 736 FSGFLVYP 743

## RESULT 12

A60032  
cerebellin-like glycoprotein - rat  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 03-Mar-1993 #sequence\_revision 03-Mar-1993 #text\_change 31-Mar-2000  
C/Accession: A60032  
R/Mada, C.; Ohtani, H.  
Brain Res. Mol. Brain Res. 9, 71-77, 1991  
A/Title: Molecular cloning of rat cerebellin-like protein cDNA which encodes a novel mem  
A/Reference number: A60032; MUID:91203483; PMID:1850079  
A/Accession: A60032  
A/Status: not compared with conceptual translation  
A/Molecule type: mRNA  
A/Residues: 1-224 <MAD>  
C/Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal hom  
C/Keywords: glycoprotein; transmembrane protein  
F/32-49/Domain: transmembrane #status predicted <TM>  
F/50-224/Domain: extracellular #status predicted <EXT>  
F/94-223/Domain: complement C1q carboxyl-terminal homology <C1Q>  
F/53,110/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 18.2%; Score 163; DB 2; Length 224;  
Best Local Similarity 38.0%; Pred. No. 3.3e-08;  
Matches 52; Conservative 15; Mismatches 62; Indels 8; Gaps 4;

QY 19 AFSARTSLVSGDAGPGRHQPLAFDTEFNIGDPPDAAGVFRCLPGAYFFSFTLGK 78  
|||:|  
Db 94 AFSR--TSTNHEBSEMNRTMTYFDQVLNIGNHFPLASSIFVAPRKGYSPFHYVK 151

QY 79 L-PRKTLISVKLMKNDDEVQAMITDDGASRRREMOSQVMLALRGDAVWLISHDHGKGA 137  
|||:|  
Db 152 VYNNQTIQVSLMONGYPIYSAFAGD-QVTRBAASNGVLLMEREDKYLKLERGNLMGG 210

QY 138 YSNHGKITTFSGFLVYP 154  
|||:|  
Db 211 W---KYSTFGSLVFP 223

## RESULT 13

S13301  
collagen alpha 1(X) chain precursor - bovine

C/Species: Bos primigenius taurus (cattle)  
C/Date: 21-Nov-1993 #sequence\_revision 23-Feb-1996 #text\_change 13-Aug-1999  
C/Accession: S13301

R/Thomas, J.T.; Kwan, A.P.L.; Grant, M.E.; Boot-Handford, R.P.  
Biochem. J. 273, 141-148, 1991

A/Title: Isolation of cDNAs encoding the complete sequence of bovine type X collagen. Ex  
A/Reference number: S13301; MUID:91113131; PMID:1703407  
A/Accession: S13301

A/Molecule type: mRNA  
A/Residues: 1-674 <THO>  
A/Cross-references: EMBL:X53556; NID:g263; PIDN:CAA37624.1; PID:g264

C/Genetics:  
A/Gene: COL10A1  
C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

C/Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer  
F/1-18/Domain: signal sequence #status predicted <SIG>  
F/19-674/Product: collagen alpha 1(X) chain #status predicted <MAT>  
F/547-673/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 18.2%; Score 163; DB 2; Length 674;  
Best Local Similarity 27.0%; Pred. No. 1.3e-07;  
Matches 51; Conservative 16; Mismatches 74; Indels 48; Gaps 4;

QY 9 GPPAPPEPSAARTSLVSGDAGPGRHQ-----  
|||:|  
Db 490 GPPGPPGPR---GNAGEPLPGPPPPGPPGVALPEDEVKAGQRFVSNAGVTGMPVS 546

QY 41 -----PLAFDTEFNIGDPPDAAGVFRCLPGAYFFSFTLGKLPKTLIS 85  
|||:|  
Db 547 AFTVLISKAYPAIGTIPFDKILYNKQOHVDPRTGIFTCKIPGIYFYSYHI-HVKGTAM 605

QY 86 YKLMKNDDEVQAMITDDGASRRREMOSQVMLALRGDAVWLISHDHGKGAYSNKGXYI 145  
|||:|  
Db 606 VGLYKNGTTPWM-YTYDEYIKGYLDQASGSAVIDLTENDQVWLPLPAGSNGLYSPRYVHS 664

QY 146 TFGFLVYP 154  
|||:|  
Db 665 SFGFLVAP 673

## RESULT 14

S15435  
collagen alpha 1(VIII) chain precursor - human

C/Species: Homo sapiens (man)  
C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 13-Aug-1999  
C/Accession: S15435

R/Muragaki, Y.; Mattei, M.G.; Yamaguchi, N.; Olsen, B.R.; Nishimura, Y.  
Eur. J. Biochem. 197, 615-622, 1991

A/Title: The complete primary structure of the human alpha-1(VIII) chain and assignment  
A/Reference number: S15435; MUID:91231001; PMID:2029894  
A/Accession: S15435

A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-744 <MR>  
A/Cross-references: EMBL:X57527; NID:g30081; PIDN:CAA40748.1; PID:g30082

C/Genetics:  
A/Gene: GDB:COL8A1  
A/Cross-references: GDB:128104; OMIM:120251

A/Map position: 3q11.1-q13.2  
C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
F/1-20/Domain: signal sequence #status predicted <SIG>  
F/21-744/Product: collagen alpha 1(VIII) chain #status predicted <MAT>

F/21-117/Region: amino-terminal nonhelical  
F/118-571/Region: interrupted helical  
F/572-743/Domain: complement C1q carboxyl-terminal nonhelical

Query Match 18.0%; Score 160.5; DB 2; Length 744;  
Best Local Similarity 25.7%; Pred. No. 2.5e-07;  
Matches 47; Conservative 26; Mismatches 71; Indels 39; Gaps 6;

QY 9 GPPAPPEPSAARTSLVSGDAGPGRHQPLAFDTEFNIGDPPDAAGVFRCLPGAYFFSFTLGKLPKTLIS 47  
|||:|  
Db 563 GPPGPPBPAPVMPRTPPPGGYLPDMGLGIDGVKPRHATGAKKGGSPAYEMPAFTAEL 622

QY 48 ---FVNIGG-----DPDAAGVFRCLPGAYFFSFTLGKLPKTLISVKLMKN 91  
|||:|  
Db 623 TAPFPVGGPVKFNKLYNGRQNYNPQTGFTCEVPGVYPAHV-HCKGQNVWVALEKN 681

QY 92 RDEVQAMITDDGASRRREMOSQVMLALRGDAVWLISHDHGKGAYSNKGKITTFSGFL 151  
|||:|  
Db 682 NEPWV-YTYDEYKKGFLDQASGSAVLLRPEDRVFLQMPSEQAAGLYAGQVHSSFSGYL 740

QY 152 VYP 154  
|||:|  
Db 741 LYP 743

## RESULT 15

S29328

complement subcomponent C1q chain C - mouse  
C/Species: Mus musculus (house mouse)  
C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 20-Aug-1999  
C/Accession: S29328

R/Petry, F.; Reid, K.B.M.; Loos, M.  
Eur. J. Biochem. 209, 129-134, 1992

A/Title: Isolation, sequence analysis and characterization of cDNA clones coding for th  
e cerebellin.  
A/Reference number: S29328; MUID:93011118; PMID:1396691

A/Accession: S29328

A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-246 <PRT>  
A:Cross-references: EMBL:X66295; NID:950228; PID:CAA46993.1; PID:950229  
C:Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal hom  
F:122-245/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 17.4%; Score 155.5; DB 2; Length 246;  
Best Local Similarity 24.4%; Pred. No. 1.9e-07;  
Matches 44; Conservative 26; Mismatches 69; Indels 41; Gaps 4;  
QY 8 RGPAPPEPRSAFSAARTSLVGS D A G P G P R H Q P----- 41  
DB 76 KGEPCMGHRGKNGRGRTSGLPD--PEPRGPGEPPGVGGRYKQKHOSVFTVTRQTOY 132  
QY 42 -----LAFDTEFVNIGSDFDAAGVFRGRLPGAYFFSFTLGKLPKRTLSVKLMKRDDEV 95  
DB 133 PEANALVRFNSTVNTPOGHYNSTGKFTCEVPGLYFYV---YTSHTANLCVHLNLNLA 188  
QY 96 QAMTYDDGASRRREMOSVMLALRRGDAVWLSDHDGCGAYSNHGKITYTSGFLVYPD 155  
DB 189 RVASFCDHMFNSKQVSSGGLRLRQGDVWLSVNDYN--GWWGIEGNSVSFGFLFPD 246

Search completed: January 12, 2004, 08:16:46  
Job time : 11.8403 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:09 ; Search time 6.10778 Seconds

(without alignments)  
1308.910 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_160\_358

Perfect score: 894

Sequence: 1 ADADAPARGPAPPPRRSAF.....LVPDLAPAAPGLGASSELL 170

Scoring table: BLOSUM62

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	894	100.0	329	1 CQT4_HUMAN	Q9bxj3 homo sapien
2	208.5	23.3	243	1 COTS_HUMAN	Q9bxj0 homo sapien
3	202.5	22.7	281	1 COT1_HUMAN	Q9bxj1 homo sapien
4	194	21.7	278	1 CQTe_HUMAN	Q9bxj9 homo sapien
5	192.5	21.5	258	1 C1RF_MOUSE	O88992 mus musculu
6	189.5	21.2	258	1 C1RF_MOUSE	O75973 homo sapien
7	188	21.0	247	1 APM1_MOUSE	O60994 mus musculu
8	187.5	21.0	285	1 CQT2_HUMAN	Q9bxj5 homo sapien
9	185.5	20.7	255	1 GLIC_MOUSE	O9esx4 mus musculu
10	182	20.4	251	1 C1QB_HUMAN	P02746 homo sapien
11	177.5	19.9	215	1 HP25_TAMSI	O06576 tamias sibi
12	175	19.6	244	1 APM1_MOUSE	O15848 homo sapien
13	171	19.1	419	1 COLE_LEPMA	P98085 leporinis mac
14	170	19.0	245	1 C1QC_HUMAN	P02747 homo sapien
15	166.5	18.6	245	1 C1QB_HUMAN	P08125 gallus galli
16	165.5	18.5	674	1 C1QB_RAT	P31721 rattus norv
17	164	18.3	680	1 C1QA_HUMAN	O03692 homo sapien
18	164	18.3	680	1 C1QA_MOUSE	O05362 mus musculu
19	163.5	18.3	744	1 C1A8_RABIT	P14282 cryptolagus
20	163	18.2	224	1 CERL_RAT	P98087 rattus norv
21	163	18.2	674	1 C1A8_BOVIN	P23206 bos taurus
22	162	18.1	744	1 C1A8_HUMAN	P27658 homo sapien
23	159	17.8	246	1 C1C1_HUMAN	Q9bxj4 homo sapien
24	155.5	17.4	246	1 C1C1_MOUSE	O02105 mus musculu
25	155.5	17.4	743	1 C1A8_MOUSE	O00780 mus musculu
26	153	17.1	193	1 CERB_HUMAN	P23435 homo sapien
27	153	17.1	193	1 CERB_MOUSE	O91171 mus musculu
28	150.5	16.8	253	1 C1QB_MOUSE	P14106 mus musculu
29	147.5	16.5	508	1 OTOT_ONCKE	P83371 oncorhynch
30	147	16.4	289	1 COT7_HUMAN	Q9bxj2 homo sapien
31	144	16.1	215	1 HP27_TAMSI	O06577 tamias sibi
32	143	16.0	196	1 HP20_TAMSI	O06575 tamias sibi
33	135.5	15.2	245	1 C1QA_HUMAN	P02745 homo sapien

34	134	15.0	245	1 C1QA_MOUSE	P98086 mus musculu
35	132.5	14.8	635	1 CA28_HUMAN	P25067 homo sapien
36	129.5	14.5	170	1 CA28_MOUSE	P25318 mus musculu
37	126.5	14.1	201	1 CERL_HUMAN	O9ntu7 homo sapien
38	89.5	10.0	265	1 Y176_HUMAN	O14681 homo sapien
39	89	10.0	384	1 DUS9_HUMAN	O99956 homo sapien
40	85	9.5	1228	1 ECV_HUMAN	O13201 homo sapien
41	80	8.9	657	1 YUXI_BACSU	P39839 bacillus su
42	77.5	8.7	664	1 KCC8_HUMAN	O13554 homo sapien
43	76.5	8.6	706	1 TRFE_HORSE	P27425 equus cabal
44	76.5	8.6	1286	1 SMC4_MOUSE	O8C647 mus musculu
45	75.5	8.4	493	1 GATR_RICPR	O9ze10 rickettsia

## ALIGNMENTS

## RESULT 1

ID CQT4\_HUMAN STANDARD; PRT; 329 AA.

AC Q9BXJ3;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DE Complement-c1q tumor necrosis factor-related protein 4 precursor.

GN C1QTNF4 OR CTRP4.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Holloway J.L., Lok S.;

RT "Homo sapiens complement-c1q tumor necrosis factor-related protein.",

Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

CC -1- SIMILARITY: Contains 2 C1Q domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----

DR EMBL: AF329838; AK17962.1;

DR Genew; HGNC:14346; C1QTNF4.

DR InterPro; IPR001073; C1q.

DR Pfam; PF00386; C1q; 2.

DR SMART; SMO0110; C1q; 2.

DR PROSITE; PS01113; C1q; 2.

DR Repeat; Signal.

FT SIGNAL 1 16

FT CHAIN 17 329

FT DOMAIN 23 159

FT FT DOMAIN 170 314

SO SEQUENCE 329 AA, 35265 MW, 331C7DBF26036915 CRC64;

Query Match 100.0%; Score 894; DB 1; Length 329;

Best Local Similarity 100.0%; Pred. No. 1,le-76;

Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ADADAPARGPAPPPRRSAFSAARTRSIVGSPAGCPRHQPLAFTPEFVNIGCDPDAAG	60
DB	160	ADADAPARGPAPPPRRSAFSAARTRSIVGSPAGCPRHQPLAFTPEFVNIGCDPDAAG	219
QY	61	VERCRLLPGAYFPFSFTLGLPKRTLSVKLMKNDEVOAMTYDDGASRRRMOQSVMALR	120
DB	220	VERCRLLPGAYFPFSFTLGLPKRTLSVKLMKNRDEVOAMTYDDGASRRRMOQSVMALR	279
QY	121	RGDAYVLLSHDDGSGVAYSNHGKVTTFSGFLVYPLDAPAPGLGASSELL	170

DB 280 RGDVWMLSHDHDGCGAYSNHGKTYTFSSGLVYVYDPLAPAPGGLGASELL 329

RESULT 2  
COT5\_HUMAN STANDARD; PRT; 243 AA.  
ID COT5\_HUMAN  
AC Q9BXJ7; Q9UXF4;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Complement-c1q tumor necrosis factor-related protein 5 precursor.  
GN C1QTNF5 OR CTRP5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sheppard P.O., Humes J.M.;  
RT "Homo sapiens complement-c1q tumor necrosis factor-related protein.";  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 25-243 FROM N.A.  
RA TISSUE=Uterus;  
RC Otsenmaelder B., Obermaier B., Mewes H.-W., Gassenhuber J.,  
RA Wiemann S.;  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: Contains 1 collagenous domain.  
CC -1- SIMILARITY: Contains 1 C1Q domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC  
CC EMBL; AF132984.1; AAK17965.1; -  
CC EMBL; AL110261; CAB53702.1; -  
CC PIR; T14782; T14782.  
CC DR GENE; HGNC:14344; C1QTNF5.  
CC DR InterPro; IPR001073; C1q.  
CC DR InterPro; IPR000087; Collagen.  
CC DR Pfam; PF00386; C1q; 1.  
CC DR Pfam; PF01391; Collagen; 1.  
CC DR PRINTS; PR00007; COMPLEMENTC1Q.  
CC DR SMART; SM00110; C1Q; 1.  
CC DR PROSITE; PS01113; C1Q; FALSE\_NEG.  
CC  
CC Collagen; Signal.  
CC  
CC FT CHAIN 16 243 POTENTIAL.  
FT SIGNAL 1 243 COMPLEMENT-C1Q TUMOR NECROSIS FACTOR-  
FT CHAIN 16 243 RELATED PROTEIN 5.  
FT DOMAIN 30 95 COLLAGEN-LIKE.  
FT FT 97 243 C1Q.  
FT FT 243 243 C1Q.  
SQ SEQUENCE 243 AA; 25298 MW; 7CCDA65CDA7EB784 CRC64;

Query Match 23.3%; Score 208.5; DB 1; Length 243;  
Best Local Similarity 35.7%; Pred. No. 1.7e-12;  
Matches 56; Conservative 18; Mismatches 72; Indels 11; Gaps 3;

DB 4 DAPAGPPAPP-----EPSSAASARTSLVSDAGPGRHPLAFTDFNVNIGDPPAA 58  
DB 85 EAGPAGPTPAGCEGSPVPSAFSAKRSERV-----PPSDAPLPDRLVNEQGHYDAV 139  
DB 59 AGVFCRLRGAYFFSTLTGLKLPKTLTVLTKMKRDEVOAMITVDDGASRRERQSQSVMLA 118  
DB 140 TKKFTQVPGVYFFA-VHATVTRASLQDFLVKNKGESIASFPQFGWPKPALSGLGAWR 198  
DB 119 LRRGDAVWMLSHDHDGCGAYSNHGKTYTFSSGLVYVYDPLAPAPGGLGASELL 155  
DB 199 LEPEDQVWVQVGVGDYIGIYASIKIDSTFTSGFLVYSD 235

RESULT 3  
COT1\_HUMAN STANDARD; PRT; 281 AA.  
ID COT1\_HUMAN  
AC Q9BXJ1; Q9GZF4;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Complement-c1q tumor necrosis factor-related protein 1 precursor  
DE (G protein coupled receptor interacting protein) (GIP).  
GN C1QTNF1 OR CTRP1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Innamorati G., Le Gouill C., Wang I., Birnbaumer M.;  
RT "GIP, a putative GPCR interacting protein.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Sheppard P.O.;  
RT "Homo sapiens complement-c1q tumor necrosis factor-related protein.";  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin.  
RX MEDLINE=2238257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Sapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udell T.B., Toshyuki S., Cantinici P., Prange C.J.,  
RA Brown S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs S.R.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skaleja U., Smalins D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [4]  
RP SEQUENCE OF 83-281 FROM N.A.  
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,  
RA Hociuta T., Hirooka S., Murakawa K., Takiguchi S., Kusano J., Chiba Y.,  
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Ishii S.,  
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,  
RA Yanamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,  
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,  
RA Wagaatsuma M., Takahashi-Fujii A., Oshino A., Sugiyama A., Kawakami B.,  
RA Suzuki Y., Sugano S., Nagahara K., Masuno Y., Nagai K., Isogai T.;  
RT "NBD human cDNA sequencing project.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: Contains 1 collagenous domain.  
CC -1- SIMILARITY: Contains 1 C1Q domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL; AJ272138; CAC20425.1; AL1\_INIT.



ID	COTE_HUMAN	STANDARD;	PRT;	278 AA.
AC	098X9;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Complement-c1q tumor necrosis factor-related protein 6 precursor.			
CN	C10NF6 OR C1RP6.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	SEQUENCE FROM N.A.			
RA	Piddington C.S., Sheppard P.O.;			
RT	"Homo sapiens complement-c1q tumor necrosis factor-related protein.";			
RL	Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Placenta;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Strusberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buecaw K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diachenko L., Marinova K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,			
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,			
RA	Boesk S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Faney J., Helton B., Kettelman M., Madan A., Rodrigues S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			

```

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield V.S.N., Krzywinski M.T., Skalska U., Smalins D.E.,
RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences." ;
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -1- SIMILARITY: Contains 1 collagenous domain.
CC -1- SIMILARITY: Contains 1 C1Q domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/sib-sib.ch).
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, AF139842; AAK17966.1; -
DR EMBL, BC020551; AAK20551.1; -
DR Genew; HGNC:14343; C1QTNF6.
DR InterPro; IPR001073; C1Q.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF00386; C1Q; 1.
DR Pfam; PF01391; Collagen; 1.
DR PRINTS; PR00007; COMPLEMENTC1Q.
DR PROSITE; PS01113; C1Q; FALSE_NEG.
DR KEGG; collagen; Signal.
FT SIGNAL 1 46
FT CHAIN 47 278
FT FT POTENTIAL.
FT FT COMPLEMENT-C1Q TUMOR NECROSIS FACTOR-
FT FT RELATED PROTEIN 6.
FT FT COLLAGEN-LIKE.
FT DOMAIN 97 138
FT CARBOHYD 139 278
FT CONFLICT 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 21 21 G -> V (IN REF. 2).
SQ SEQUENCE 278 AA; 30861 MW; 27A82CA863F23D47 CAC64;

Query Match 21.7%; Score 194; DB 1; Length 278;
Best Local Similarity 41.8%; Pred. No. 4.6e-11;
Matches 61; Conservative 10; Mismatches 65; Indels 10; Gaps 5.

QY 10 PPAPPEPR-SAFSAARTRSLVSGDAGQPPRPQPLAFDTPEFNIIGDGFPAAGVRCRLPG 68
DB 135 PGACQCKRFFAFSGRKRTAL-----HSGEDFQTLTFEVPFNILGSCFPMATGQFAAPRG 189
QY 69 AYPFSFTLGKLPRTLLSYKLMKNDDEVQAMTYDGAISRREMOSQSWLALRGDAVWL 128
DB 190 IYFSLNVHSMYNYETVYVHIMNKE--AVILYAQPSERSIMQSQSWLIDLAYGDVWVR 247
QY 129 SHHDGYG-A-YSN-HGKIYTFSGFLV 152
DB 248 LFKRGRENATYSNDFYITTSGLHI 273

RESULT 5
CIRF_MOUSE
ID CIRF_MOUSE STANDARD; PRT; 258 AA.
AC 088932;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE C1q-related factor precursor.
DE C1QRF OR CRF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99097006; PubMed=9878755.
RA Berube N.G., Swanson X.H., Bertiam M.J., Kittle J.D., Didenko V.,
RA Baekin D.S., Smith J.R., Pereira-Smith O.M.;
RT "Cloning and characterization of CRF, a novel C1q-related factor,
```

[illegible]

Query Match	21.5%	Score 192.5	DB 1	Length 258
Best Local Similarity	29.7%	Pred. No. 5	8e-11	
Matches	55	Conservative	20	Mismatches 67, Indels 43, Gaps 5

  

Qy	9	GPPAPPEPRSAFSAARTSLVGSADGPGPRHP	-----	41
Db	79	GPPGPPDRGPPGPGVPEKGEPPKPGPPGLPGSGSGAISTATYTPVPAFYAGLKN		138
Qy	42	-----LAFDFEPVNIIGDPPAAAGVPPRCRLPGAYFSPFTGLKLP	-----	KTLSVKLMK 90
Db	139	PHEGEYVKKFDDVVTNLGNVYDAAAGKFTCNIPGYFFTYHV	--LMRGDGTSMADLCK	196
Qy	91	NRDEYQAMVYDDGASRRRMOSSQVWLALRCDGDAWLLSHDDGCGAYSNHGKYYTFSGF		150
Db	197	N-GGVRAASIAQDADQNDYDANSYVLIHLDADDEVFIKLDGSKAKHAGNSN	--KYTFSSGF	253
Qy	151	LVYPD	155	
Db	254	IYSD	258	

  

RESULT 6			
CIRF_HUMAN			
ID_CIRF_HUMAN	STANDARD;	PRT;	258 AA.
AC	O75973;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	C1q-related factor precursor.		
GN	C1ORF OR CRF.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=9097006; PubMed=9878755;		
RA	Beribe N.G., Swanson X.H., Bertam M.J., Kittle J.D., Didenko V., Baekin D.S., Smith J.R., Pereira-Smith O.M.;		

RT "Cloning and characterization of CRF, a novel C1q-related factor,  
RL expressed in areas of the brain involved in motor function." ;  
RN Brain Res. Mol. Brain Res. 63:233-240(1999).  
RP [2]  
RP SEQUENCE FROM N.A.  
RA Rieder M.J., Carrington D.P., Chung M.-W., Lee K.L., Poiel C.L., Yi Q.,  
RA Nickerson D.A. ;  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RP TISSUE=Placenta ;  
RC MEDLINE=22388257 ; PubMed=12477932 ;  
RA Strausberg R.L., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,  
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,  
RA Aleschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Dackiwko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Steplento M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ueslin T.B., Toshitsuki S., Caraminci P., Prange C.,  
RA Raha S.S., Loguallano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Pahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butcherfield Y.S.N., Krzywiński M.I., Skelton S., Smalhus D.E.,  
RA Schermer A., Schein J.E., Jones S.J.W., Marra M.A. ;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences." ;  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAINSTEM.  
CC -1- SIMILARITY: Contains 1 collagenous domain.

```

CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC      or send an email to license@ebi.ac.uk).
CC      -----
DR      EMBL; AF095154; AAC6148.1; -
DR      EMBL; AF410771; AAK95248.1; -
DR      EMBL; BC008798; AAH08798.1; -
DR      GO; GO:00076526; P:locomotory behavior; NAS.
DR      InterPro; IPR001073; C1q.
DR      InterPro; IPR000087; Collagen.
DR      Pfam; PF00386; C1q; 1.
DR      Pfam; PF01391; Collagen; 1.
DR      PRINTS; PR00007; COMPLEMENTC1Q.
DR      SMART; SM00110; C1Q; 1.
DR      ProSITE; PS01113; C1Q; 1.
DR      Collagen; Signal.
KW      SIGNAL
FT      CHAIN          1         16      POTENTIAL.
FT      CHAIN          17        258      C1Q-RELATED FACTOR.
FT      DOMAIN         67        115      COLLAGEN-LIKE.
FT      DOMAIN         123        258      C1Q.
SQ      SEQUENCE      258 AA;  26452 MW;  52C51CDF59C4E2BF CRC64;

Query Match          21.2%;  Score 189.5;  DB 1;  Length 258;
Best Local Similarity 32.0%;  Pred.No. 1.1e-10;
Matches      54;  Conservative      22;  Mismatches      58;  Indels      35;  Gaps      6

QY      9  GPPAP-----EPRNAPSAAKRRSLVGSAGGCGPRHQLAPDTEFVN 50
DB      103 GPPGPPGPPGAGGSAISTATYTVPRVAFYAGL-----KNHGEYEVAKFDDVVTN 154

QY      51 IGGDDNAGAVRCRLPGAYPFSTLGLPR-----KTLSEVKLMKRDEVOAMIVDDGASR 106
DB      155 LGNNNDASGKFTCNIGTGYFTTHAV--LMRGSGDGTSMNADLCKG-GGVRSALINODADQ 211

```

107 REMOSOSVMTALRGDAVWMLSHDHDCGAVSNHGKXITTSGLVYPD 155  
 Db 212 NYDVASNSVILHLADGDEVFLTKLDGKAHGGNSN--KXSTFSGLIYVD 258  
 RESULT 7  
 APML\_MOUSE STANDARD: PRT: 247 AA.  
 ID APML\_MOUSE Q62400; Q9DC68;  
 AC Q60994; Q62400; Q9DC68;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Adiponectin precursor (30 kDa adipocyte complement-related protein)  
 GN ACP30 (Adipocyte specific protein AdipoQ).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Adipocyte;  
 RX MEDLINE=9620757; PubMed=7592907;  
 RA Scherer P.E., Williams S., Fogliano M., Baldini G., Lodish H.F.;  
 RT "A novel serum protein similar to C1q, produced exclusively in  
 adipocytes.";  
 RL J. Biol. Chem. 270:26746-26749(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Fibroblast;  
 RX MEDLINE=9620999; PubMed=8631877;  
 RA Hu B., Liang P., Spiegelman B.M.;  
 RT "AdipoQ is a novel adipose-specific gene dysregulated in obesity.";  
 RL J. Biol. Chem. 271:10697-10703(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=11162643;  
 RA Das K., Lin Y., Widen E., Zhang Y., Scherer P.E.;  
 RT "Chromosomal localization, expression pattern, and promoter analysis  
 of the mouse gene encoding adipocyte-specific secretory protein  
 Acip30.";  
 RL Biochem. Biophys. Res. Commun. 280:1120-1125(2001).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Heart;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 Arkawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 Aitawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 Kadota K., Matsuda H.A., Ashburner M., Batilov S., Caavaant T.,  
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,  
 Kuehl P., Lewis S., Matsumoto Y., Nakai I., Pesole G., Quackenbush J.,  
 Schirral L.M., Steudli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 Gustingich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
 Lyons P., Marchionni L., Mashima J., Mazzarelli I., Mombereis P.,  
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,  
 Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 Wyshew-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlski S.,  
 Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 403:685-690(2001).  
 RN [5]  
 RP FUNCTION.  
 RX MEDLINE=21372498; PubMed=11479627;  
 RA Yamaguchi T., Kamon J., Waki H., Terachi Y., Kubota N., Hara K.,  
 Mori Y., Ide T., Murakami K., Tsuboyama-Kaseoka N., Ezaki O.,  
 Akanuma Y., Gavrilova O., Vinson C., Reitman M.L., Kagechika H.,

RA Shudo K., Yoda M., Nakano Y., Tobe K., Nagai R., Kimura S., Tomita M.,  
 RA Froguel P., Kadowaki T.;  
 RT "The fat-derived hormone adiponectin reverses insulin resistance  
 associated with both lipodystrophy and obesity.";  
 RL Nat. Med. 7:941-946(2001).  
 RN [6]  
 RP FUNCTION.  
 RX MEDLINE=21372499; PubMed=11479628;  
 RA Berg A.H., Combs T.P., Du X., Brownlee M., Scherer P.E.;  
 RT "The adipocyte-secreted protein Acip30 enhances hepatic insulin  
 action.";  
 RL Nat. Med. 7:947-953(2001).  
 CC -1- FUNCTION: IMPORTANT NEGATIVE REGULATOR IN HEMATOPOIETIC AND IMMUNE  
 SYSTEMS; MAY BE INVOLVED IN ENDING INFLAMMATORY RESPONSES THROUGH  
 ITS INHIBITORY FUNCTIONS. INHIBITS ENDOTHELIAL NF-KAPPA-B SIGNALING  
 THROUGH A CAMP-DEPENDENT PATHWAY. INHIBITS TNF-ALPHA-INDUCED  
 EXPRESSION OF ENDOTHELIAL ADHESION MOLECULES. INVOLVED IN THE  
 CONTROL OF FAT METABOLISM AND INSULIN SENSITIVITY.  
 CC -1- SUBUNIT: Homodimer.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Synthesized exclusively by adipocytes and  
 secreted into plasma.  
 CC -1- INDUCTION: DURING HORMONE-INDUCED ADIPOSE DIFFERENTIATION AND  
 ACTIVATED BY INSULIN.  
 CC -1- SIMILARITY: Contains 1 C1Q domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL: U37222; AAA80543.1; -;  
 DR EMBL: U49915; AAB06706.1; -;  
 DR EMBL: AF304466; AAK13417.1; -;  
 DR EMBL: AK003138; BAB22597.1; -;  
 DR PDB: 1C28; 07-SEP-99.  
 DR MGD: MGI:106675; Acip30.  
 DR GO: GO:0005576; C:extracellular; IDA.  
 DR GO: GO:0005515; F:protein binding activity; IMP.  
 DR GO: GO:0006635; P:fatty acid beta-oxidation; IMP.  
 DR InterPro: IPR001073; C1Q.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF00336; C1Q; 1.  
 DR Pfam: PF01391; Collagen; 1.  
 DR PRINTS: PR00007; COMPLEMENTC1Q.  
 DR PRODOM: PD000007; C1Q helix; 1.  
 DR SMART: SM0110; C1Q; 1.  
 DR PROSITE: PS0113; C1Q; 1.  
 KW Hormone; Collagen; Signal; Repeat; Hydroxylation; Plasma;  
 KW Polymorphism; 3D-structure.  
 KM SIGNAL 1  
 FT CHAIN 17  
 FT SIGNAL 18  
 FT CHAIN 247  
 FT DOMAIN 45 110  
 FT DOMAIN 111 247  
 FT DISULFID 39 39  
 FT MOD\_RES 47 47  
 FT MOD\_RES 50 50  
 FT MOD\_RES 56 56  
 FT MOD\_RES 65 65  
 FT MOD\_RES 79 79  
 FT MOD\_RES 98 98  
 FT MOD\_RES 107 107  
 FT MOD\_RES 113 113  
 FT VARIANT 50 50  
 FT CONFLICT 74 74  
 FT CONFLICT 117 117  
 FT CONFLICT 148 148  
 FT CONFLICT 243 243  
 FT SEQUENCE 247 AA; 26841 MW; 137B687D873988C4 CRC64;

Query Match 21.0%; Score 188; DB 1; Length 247;  
 Best Local Similarity 33.5%; Pred. No. 1.5e-10;  
 Matches 53; Conservative 24; Mismatches 73; Indels 8; Gaps 5;

QY 3 ADAPRAGPAPP-----EPSSASARTSLVSDAGPGRHQPPLAFDFEVNIGDPPDA 58  
 DB 91 AEGP-RGFPPTGRKGEPEEAAYMYRSAPSVGLETRVTVPNVPIRFTKIFVYQNHQDGS 149  
 QY 59 AGVPRCRLPGAYFFSFTLGKLPKRTLSVKLMKNRDEVQAMTYDDCASRRREMOSQSVMLA 118  
 DB 150 TCKFTCNIPGLYFYFVYHI-TVYMKDVKYSLFK-KKAVLFTTDYQEKXVDQASGVLLH 207  
 QY 119 LRGDVAVMLSH-DHDGCVAYSNHGKTYTFSGFLVYPP 155  
 DB 208 LEVGDQVWLVQYVGGDGHNGLVADVNDVSTFTGFLLYHD 245

RESULT 8  
 COT2\_HUMAN STANDARD; PRT; 285 AA.

ID COT2\_HUMAN  
 AC Q9BXJ5;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Complement-c1q tumor necrosis factor-related protein 2 precursor.  
 GN C1QTNF2 OR CTRP2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RA Piddington C.S., Bishop P.;  
 RT "Homo sapiens complement-c1q tumor necrosis factor-related protein.";  
 RL Submitted (DEC-2000) to the EMBL/Genbank/DDBJ databases.  
 RN 12  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Muscle;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strusberg R.L., Feingold E.A., Georger L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
 RA Altchut S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stalcenon M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shcherchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,  
 RA Butcherfield V.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- SIMILARITY: Contains 1 collagenous domain.  
 CC -1- SIMILARITY: Contains 1 C1Q domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AF329836; AK17960.1; -;  
 DR EMBL: BC011699; AA11699.1; -;

DR Genew; HGNC:14325; C1QTNF2.  
 DR InterPro; IPR001073; C1q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00385; C1q; 1.  
 DR Pfam; PF01391; Collagen; 2.  
 DR PRINTS; PRO0007; COMPLEMENTC1Q.  
 DR SMART; SM01110; C1Q; 1.  
 DR PROSITE; PS01113; C1Q; 1.  
 KW Collagen; Signal.  
 FT SIGNAL 1 15  
 FT CHAIN 16 285  
 FT DOMAIN 40 141  
 FT DOMAIN 143 285  
 FT SEQUENCE 285 AA; 29952 MW; 7E31FF9868D4EDFA CRC64;  
 POTENTIAL.  
 COMPLEMENT-C1Q TUMOR NECROSIS FACTOR-  
 RELATED PROTEIN 2.  
 COLLAGEN-LIKE.  
 C1Q.

Query Match 21.0%; Score 187.5; DB 1; Length 285;  
 Best Local Similarity 32.5%; Pred. No. 1.9e-10;  
 Matches 52; Conservative 31; Mismatches 52; Indels 25; Gaps 8;

QY 8 RGPAPPEP-----RSAPSAARTSLVSDAGPGRHQPPLAFDFEVNIGDPPDA 58  
 DB 132 KGEPLPGPCSCSGHTRKSAFSAVATKSY-----PRERLPIKPKILNNEGHNAS 183  
 QY 59 AGVPRCRLPGAYFFSFTLGKLPKRTLSVKLMKNRDEVQAMTYDDCASRRREMOSQSVMLA 118  
 DB 184 SGKEFVGVPGLIYFTYDI-TLANKGLAIGLVHN-GQYRIRTFDANTG-NHDVAGSSTLA 240  
 QY 119 LRGDVAVMLSHDHDGCVAYSNHGKTYTFSGFLVYPP 155  
 DB 241 LKQGEVWLVQYVGGDGHNGLVADVNDVSTFTGFLLYAD 278

RESULT 9  
 GLIC\_MOUSE STANDARD; PRT; 255 AA.

ID GLIC\_MOUSE  
 AC Q9BSN4;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Gliacolin precursor (C1q-like protein).  
 DE Gliacolin precursor (C1q-like protein).  
 GN C1QD.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCB1\_TaxID=10090;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=20428709; PubMed=10862615;  
 RA Koide T., Aso A., Yoshizumi T., Nagata K.;  
 RT "Conformational requirements of collagenous peptides for recognition  
 RT by the chaperone protein HSP47.";  
 RL J. Biol. Chem. 275:27957-27963(2000).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN GLIAL CELLS.  
 CC -1- SIMILARITY: Contains 1 collagenous domain.  
 CC -1- SIMILARITY: Contains 1 C1Q domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AB044560; BAB15806.1; -;  
 DR MGI; MGI:2387350; C1q1.  
 DR GO; GO:0005515; F:protein binding activity; IPT.  
 DR InterPro; IPR001073; C1q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1q; 1.  
 DR Pfam; PF01391; Collagen; 1.  
 DR PRINTS; PRO0007; COMPLEMENTC1Q.

DR SMART; SM00110; C1Q; 1.  
 DR PROSITE; PS01113; C1Q; 1.  
 KW Collagen; Signal.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 255 GLIACOLIN.  
 FT DOMAIN 61 111 COLLAGEN-LIKE.  
 FT DOMAIN 120 255 C1Q.  
 SQ SEQUENCE 255 AA; 26687 MW; 529FBAFB2191BC1 CRC64;  
 Query Match 20.7%; Score 185.5; DB 1; Length 255;  
 Best Local Similarity 28.0%; Pred. No. 2.6e-10;  
 Matches 53; Conservativity 21; Mismatches 72; Indels 43; Gaps 5;  
 QY 5 APARGPAPPEPSAFAATRSIVSGDAPGPRHQP----- 41  
 DB 72 AGRGPGGPGGPGVPGPGKEKPGQGPGPGAGLNAAGASATYSTVPKIAFYA 131  
 QY 42 -----LAFDFEVNIGGDFDAAGVPRCRLPGAYFSPFTLGKLP----- KTLAV 86  
 DB 132 GLKRGHGVVLFKFDVYVTLGNHYDPTTKFTCSIPGIYFFTYHV--LMRGDGTSMWA 189  
 QY 87 KLMKNDEVQAMTYDDGASHRRMOSQVLMALRRGDAVWLSDHDGVGYVSNHGKXYIT 146  
 DB 190 DLCKN-NOVRAASAIADADQNDYASNSVYLHLEPGDEVYIKLDGKGKAG--GNNNKYST 246  
 QY 147 FSGFLVYYPD 155  
 DB 247 FSGFIIYAD 255  
 RESULT 10  
 C1QB\_HUMAN STANDARD; PRT; 251 AA.  
 ID P02746; Q96H17;  
 AC 21-JUL-1986 (Rel. 01, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Complement C1q subcomponent, B chain precursor.  
 GN C1QB.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86076906; PubMed=3000358;  
 RA Reid K.B.M.;  
 RT "Molecular cloning and characterization of the complementary DNA and  
 gene coding for the B-chain of subcomponent C1q of the human  
 complement system.";  
 RT Biochem. J. 231:729-735 (1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 Stapleton M., Soares M.B., Bonaldo W.F., Casavant T.L., Scheetz T.E.,  
 Brownstein M.U., Ussid T.B., Toshiyuki S., Carninci P., Prange C.,  
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 Fahy J., Heiton E., Kettelman M., Madan A., Rodriguez A.,  
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 Blakeley R.W., Tuchman J.W., Green E.D., Dickson M.C.,  
 Rodriguez A.C., Grumman J., Schmutz J., Myers R.M.,  
 Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,  
 Scherch A., Schein J.E., Jones S.J.M., Marris M.A.,  
 "Generation and initial analysis of more than 15,000 full-length

RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [3]  
 RP SEQUENCE OF 26-133.  
 RX MEDLINE=80020137; PubMed=486087;  
 RA Reid K.B.M.;  
 RT "Complete amino acid sequences of the three collagen-like regions  
 present in subcomponent C1q of the first component of human  
 complement.";  
 RT Biochem. J. 179:367-371 (1979).  
 RN [4]  
 RP SEQUENCE OF 26-193.  
 RX MEDLINE=79041552; PubMed=708376;  
 RA Reid K.B.M., Thompson E.O.P.;  
 RT "Amino acid sequence of the N-terminal 108 amino acid residues of the  
 B chain of subcomponent C1q of the first component of human  
 complement.";  
 RT Biochem. J. 173:863-868 (1978).  
 RN [5]  
 RP SEQUENCE OF 134-251.  
 RX MEDLINE=82283890; PubMed=6981411;  
 RA Reid K.B.M., Gagnon J., Frampton J.;  
 RT "Completion of the amino acid sequences of the A and B chains of  
 subcomponent C1q of the first component of human complement.";  
 RL Biochem. J. 203:559-569 (1982).  
 RN [6]  
 RP SEQUENCE OF 224-251 FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=85038855; PubMed=6208566;  
 RA Reid K.B.M., Bentley D.R., Wood K.J.;  
 RT "Cloning and characterization of the complementary DNA for the B  
 chain of normal human serum C1q.";  
 RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 306:345-354 (1984).  
 RN [7]  
 RP REVIEW OF C1Q DEFICIENCY.  
 RX MEDLINE=98450587; PubMed=9777412;  
 RA Petry F.;  
 RT "Molecular basis of hereditary C1q deficiency.";  
 RL Immunobiology 199:286-294 (1998).  
 CC [1] FUNCTION: C1Q ASSOCIATES WITH THE PROENZYMES C1R AND C1S TO YIELD  
 CC C1. THE FIRST COMPONENT OF THE SERUM COMPLEMENT SYSTEM. THE  
 CC COLLAGEN-LIKE REGIONS OF C1Q INTERACT WITH THE CA(12+)-DEPENDENT  
 CC TAK(2)CIS(2) PROENZYME COMPLEX, AND EFFICIENT ACTIVATION OF C1  
 CC TAKES PLACE ON INTERACTION OF THE GLOBULAR HEADS OF C1Q WITH THE  
 CC FC REGIONS OF IGG OR IGM ANTIBODY PRESENT IN IMMUNE COMPLEXES.  
 CC [2] SUBUNIT: C1 IS A CALCIUM-DEPENDENT TRIMOLECULAR COMPLEX OF C1Q, R  
 CC AND S IN THE MOLAR RATION OF 1:2:2. C1Q SUBCOMPONENT IS COMPOSED  
 CC OF NINE SUBUNITS, SIX OF WHICH ARE DISULFIDE-LINKED DIMERS OF THE  
 CC A AND B CHAINS, AND THREE OF WHICH ARE DISULFIDE-LINKED DIMERS OF  
 CC THE C CHAIN.  
 CC [3] PTM: O-linked glycans consist of Glc-Gal disaccharides bound to  
 CC the oxygen atom of post-translationally added hydroxyl groups.  
 CC [4] DISEASE: Defects in C1QB are a cause of C1Q deficiency  
 CC [MIM:120570]. It is a rare genetic disorder which is associated  
 CC with recurrent infections and a high prevalence of lupus  
 CC erythematosus-like symptoms. It is characterized by a loss of  
 CC activation of the complement classical pathway.  
 CC [5] SIMILARITY: Contains 1 collagenous domain.  
 CC [6] SIMILARITY: Contains 1 C1Q domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 DR EMBL; X03084; CA26880.1; -  
 DR EMBL; BC008983; AA041692.1; -  
 DR EMBL; M36278; AAC41692.1; -  
 DR Gene; HGNC:1242; C1QB.  
 DR MIM; 120570; -

```

CC Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
CC Tamias
OX NCBI_TaxID=64680;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RC MEDLINE=93180798; Pubmed=8441393;
RA Takamatsu N., Ohba K., Kondo J., Kondo N., Shiba T.;
RT "Hibernation-associated gene regulation of plasma proteins with a
RT collagen-like domain in mammalian hibernators.";
RN Mol. Cell. Biol. 13:1516-1521(1993).
[2]
RN SEQUENCE OF 29-62; 84-130; 172-183; 187-192 AND 201-215.
RC TISSUE=Plasma;
RC MEDLINE=92112696; Pubmed=1730610;
RA Kondo N., Kondo J.;
RT "Identification of novel blood proteins specific for mammalian
RT hibernation.";
RN J. Biol. Chem. 267:473-478(1992).
CC -1- FODUCTON: PLASMA PROTEINS HP-20, HP-25, HP-27 AND HP-55 FORM A
CC 140 KDa COMPLEX VIA DISULFIDE BONDS IN THE PLASMA AND ARE
CC HIBERNATION SPECIFIC.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: PLASMA; SYNTHESIZED IN THE LIVER.
CC -1- DEVELOPMENTAL STAGE: THE PROTEIN COMPLEX DISAPPEARS FROM THE
CC PLASMA AT ONSET OF HIBERNATION AND REAPPEARS AS HIBERNATION
CC CEASES.
CC -1- SIMILARITY: Contains 1 collagenous domain.
CC -1- SIMILARITY: Contains 1 C1Q domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL, D12975; BA002352.1; .
CC PIR, B48150; B48150.
CC InterPro: IPR001073; C1Q.
CC InterPro: IPR000087; Collagen.
CC Pfam, PF00386; C1Q; 1.
CC Pfam, PF01391; Collagen; 1.
CC PRINTS, PR00007; COMPLENNTC1Q.
CC SMART, SMO0110; C1Q; 1.
CC PROSITE, PS01113; C1Q; 1.
CC Signal, collagen; Glycoprotein; Plasma; Multigene family.
KW SIGNAL
FT 1 28
FT CHAIN
FT 29 215
FT FT HIBERNATION-ASSOCIATED PLASMA PROTEIN
FT FT HP-25.
FT FT COLLAGEN-LIKE.
FT DOMAIN 40 81
FT DOMAIN 83 215
FT CARBOHYD 167 167 C1Q.
FT SEQUENCE 215 AA; 22664 MW; AFE03206917EA530 CRC64;
-----
Query Match 19.9%; Score 177.5; DB 1; Length 215;
Best Local Similarity 32.4%; Pred. No. 1.2e-09;
Matches 58; Conservative 21; Mismatches 49; Indels 51; Gaps 10
-----
QY 5 AP-ARGPAPF-----EPRAFSARTRSLVSGDAGGPPHPLA 43
DB 56 AFGALGPPPPVPGIPGPGPPGDDVEKSSPKSAFAVKL-----SERPEP-FQPIV 108
QY 44 FDFEFNIGDDEDAAGVFRCLPGAFFSFPTLGLPKRTLSVKLMKRDVQAMITYDG 103
DB 109 FKALYNGGHNMTGERSCLPGVYNGFDI-RLFGSSVKIRLM--RDGIQV----- 159
QY 104 ASRRRMOGO-----SVMLALRGDAVWLISHDHQGYGAYSNHG-KYIFSGFLVY 153
DB 160 --REKVAQNDSYKHAMGSVIVALGSGDGVWLESKLK--GTESEKGIITHIVFGYLLY 213

```

ID	APM1_HUMAN	STANDARD;	PRT;	244 AA.
AC	Q15648;			
DT	01-NOV-1997	(Rel. 35, Created)		
DT	01-NOV-1997	(Rel. 35, Last sequence update)		
DT	15-SEP-2003	(Rel. 42, Last annotation update)		
DE	Adiponectin precursor (30 kDa adipocyte complement related protein)			
DE	(ACBP30) (Adipose most abundant gene transcript 1) (APM-1) (Gelatin-binding protein).			
GN	APM1 OR ACBP30 OR GBP28.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxId=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Adipose tissue;			
RX	MEDLINE=96224171; PubMed=8619847;			
RA	Maeda K., Okubo K., Shimomura I., Funahashi T., Matsuzawa Y.,			
RA	Matsubara K.;			
RT	"cDNA cloning and expression of a novel adipose specific collagen-like			
RT	factor. APM1 (Adipose Most abundant Gene transcript 1).";			
RL	Biochem. Biophys. Res. Commun. 221:286-289(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99196984; PubMed=10095105;			
RA	Saito K., Tobe T., Minoshima S., Aakawa S., Sumiya J., Yoda M.,			
RA	Nakano Y., Shimizu N., Tomita M.;			
RT	"Organization of the gene for gelatin-binding protein (GBP28).";			
RL	Gene 229:67-73(1999).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99333693; PubMed=10403784;			
RA	Schaeffler A., Orso E., Pallitzsch K.D., Buechler C., Drobnik W.,			
RA	Fuerst A., Schoelmerich J., Schmitt G.;			
RT	"The human apm-1, an adipocyte-specific gene linked to the family of			
RT	TNF's and to genes expressed in activated T cells, is mapped to			
RT	chromosome 1q21.3-q23, a susceptibility locus identified for familial			
RL	combined hyperlipidemia (FCH).";			
RL	Biochem. Biophys. Res. Commun. 260:416-425(1999).			
RN	[4]			
RP	CHARACTERIZATION.			
RX	MEDLINE=20417747; PubMed=10961870;			
RA	Yokota T., Oritani K., Takahashi T., Ishikawa J., Matsuyama A.,			
RA	Ouchi N., Kihara S., Funahashi T., Tenner A.J., Tomiyama Y.,			
RA	Matsuzawa Y.;			
RT	"Adiponectin, a new member of the family of soluble defense collagens,			
RT	negatively regulates the growth of myelomonocytic progenitors and the			
RT	functions of macrophages.";			
RL	Blood 96:1723-1732(2000).			
RN	[5]			
RP	CHARACTERIZATION.			
RX	MEDLINE=20440368; PubMed=10982546;			
RA	Ouchi N., Kihara S., Arita Y., Okamoto Y., Maeda K., Kuriyama H.,			
RA	Hotta K., Nishida M., Takahashi M., Muraguchi M., Ohmoto Y.,			
RA	Kakimura T., Yamashita S., Funahashi T., Matsuzawa Y.;			
RT	"Adiponectin, an adipocyte-derived plasma protein, inhibits			
RT	endothelial NF-kappaB signaling through a cAMP-dependent pathway.";			
RL	Circulation 102:11296-1301(2000).			
RN	[6]			
RP	FUNCTION.			
RX	MEDLINE=21372498; PubMed=11479627;			
RA	Yamauchi T., Kamon J., Maki H., Terauchi Y., Kubota N., Hara K.,			
RA	Mori Y., Ide T., Murakami K., Tsuboyama-Kasaoka N., Ezaki O.,			
RA	Ashikawa Y., Gavrilova O., Vinson C., Reitman M.L., Kagechika H.,			
RA	Shukuna K., Yoda M., Nakano Y., Tobe K., Nagai R., Kimura S., Tomita M.,			
RA	Froguel P., Kadwaki T.;			
RT	"The fat-derived hormone adiponectin reverses insulin resistance			
RT	associated with both lipodystrophy and obesity.";			
RL	Nat. Med. 7:941-946(2001).			
RN	[7]			

FT MOD RES 76 76 HYDROXYLATION (BY SIMILARITY).  
 FT MOD RES 86 86 HYDROXYLATION (BY SIMILARITY).  
 FT MOD RES 95 95 HYDROXYLATION (BY SIMILARITY).  
 FT MOD RES 104 104 HYDROXYLATION (BY SIMILARITY).  
 FT VARIANT 84 84 G -> R.  
 FT VARIANT 112 112 /FTID=VAR\_013273.  
 FT VARIANT 112 112 R -> C (in adiponectin deficiency).  
 FT VARIANT 117 117 V -> M.  
 FT VARIANT 117 117 /FTID=VAR\_013274.  
 FT VARIANT 164 164 I -> T.  
 FT VARIANT 164 164 /FTID=VAR\_013275.  
 FT VARIANT 221 221 R -> S.  
 FT VARIANT 221 221 /FTID=VAR\_013276.  
 FT VARIANT 241 241 H -> P.  
 FT VARIANT 241 241 /FTID=VAR\_013277.  
 SQ SEQUENCE 244 AA; 26414 MW; 64DB6C1204B1018 CRC64;  
 Query Match 19.6%; Score 175; DB 1; Length 244;  
 Best Local Similarity 31.1%; Pred. No. 2.4e-09;  
 Matches 51; Conservative 24; Mismatches 65; Indels 24; Gaps 6;  
 QY 7 ARGPAPP-----EF-----RSAPSAARTSLVSGDAGPGPRHOPLAFDTEFVNIG 52  
 DB 88 AEGRGPRPGIGRKGEPEGAYVRSAPF-----VGLTYVTI PNMPIRFTKIFVNOQ 140  
 QY 53 GDDPAAAGVRCRLPGAYFSPFTLGKLPKRLTSYKLMKNDEVOAMTYDGCASRRRMOG 112  
 DB 141 NHVDGSGTKFHCHNPGLYYFAVYHITVYMKDVKSLER-KDKAMLFYDYOQENNVQAS 198  
 QY 113 OSVVALRRGDVAWLISH-DHDGYGAYSNHGKTYTFSGFLVYPD 155  
 DB 199 GSVLLHLEVGDVQLVQYVGEGERGLADNNDSTFTGFLYHYD 242  
 RESULT 13  
 COLE LEPPA STANDARD; PRT; 419 AA.  
 ID ID COLE LEPPA STANDARD; PRT; 419 AA.  
 AC P96085; Q91080; (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Inner ear-specific collagen precursor (Saccular collagen).  
 OS Lepomis macrochirus (Bluegill).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 CC Acanthomorphi; Acanthopterygii; Percomorpha; Perciformes; Percoidae;  
 CC Centrarchidae; Lepomis.  
 CX NCBI\_Taxid=13106;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95167486; PubMed=786331;  
 RA Davis J.G., Oberholzer J.C., Burns F.R., Greene M.I.;  
 RT "Molecular cloning and characterization of an inner ear-specific  
 structural protein".  
 RT Science 267:1031-1034 (1995).  
 RN [2]  
 RP CONCEPTUAL TRANSLATION.  
 RA Gibson T.;  
 RL Submitted (MAR-1995) to the SWISS-PROT data bank.  
 CC - FUNCTION: FORMS A MICROSTRUCTURAL MATRIX WITHIN THE OTOLITHIC  
 CC MEMBRANE (PROBABLY).  
 CC - TISSUE SPECIFICITY: SPECIALIZED SECRETORY SUPPORTING CELLS AT THE  
 CC - OUTER PERIMETER OF THE SACCULAR EPITHELIUM.  
 CC - SIMILARITY: Contains 1 C1Q domain.  
 CC - CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT HAD TO BE  
 CC INTRODUCED FOR POSITIONS 391-419 SO AS TO MAXIMIZE THE SIMILARITY  
 CC WITH OTHER SHORT-CHAIN COLLAGENS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U17431; AAA69978.1; ALT\_FRAME.  
 DR InterPro: IPR01073; C1q.  
 DR InterPro: IPR00087; Collagen.  
 DR Pfam: PF00386; C1q; 1.  
 DR Pfam: PF01391; Collagen; 3.  
 DR PRINTS: PRO00007; COMPLEMENTC1Q.  
 DR PRODOM: PD000007; C1q\_helix; 2.  
 DR SMART: SM00110; C1q; 1.  
 DR PROSITE: PS01113; C1q; 1.  
 KW Extracellular matrix; Repeat; Collagen; signal.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 419 INNER EAR-SPECIFIC COLLAGEN.  
 FT DOMAIN 20 57 NONHELICAL REGION (NC2).  
 FT DOMAIN 58 274 TRIPLE-HELICAL REGION (COL1).  
 FT DOMAIN 275 419 NONHELICAL REGION (NC1).  
 FT DOMAIN 272 419 C1Q.  
 FT CARBOHYD 37 37 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 320 320 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 419 AA; 43634 MW; 570CDB9675FC0F39 CRC64;  
 Query Match 19.1%; Score 171; DB 1; Length 419;  
 Best Local Similarity 31.1%; Pred. No. 1.1e-08;  
 Matches 59; Conservative 14; Mismatches 65; Indels 52; Gaps 8;  
 QY 4 DARGPAPP-----EPSSAFSAARTSLVSGS 31  
 DB 234 DTGARPGPPGGRGMAGLRGKLVGRCPKPGSPGSAVQISAPSV-----GLFSP 289  
 QY 32 DAGRPHPHPLAFDTEFVNIGDPPDAAGVRCRLPGAYFSP--TLGKLP-RKTL---S 85  
 DB 290 RSPF-PSPDPVKTKFYNEBGMPTLNKFNFTYRGVILFSYHITVRNPPVALVVG 348  
 QY 86 VKLMKNDEVQAMTYDGCASRRRMOGOSVVALRRGDVAWLISHDHDGYGAYSNHGKTY 145  
 DB 349 VRKLRTDSLXGQDIDA-----SNLALHLTDGQVWLFTL-RDMNGXVSSSDSDS 399  
 QY 146 TFGSFLVYPD 155  
 DB 400 TFGSFLVYPD 409  
 RESULT 14  
 C1Qc HUMAN STANDARD; PRT; 245 AA.  
 ID ID C1Qc HUMAN STANDARD; PRT; 245 AA.  
 AC P02747; Q96H05; (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Complement C1q subcomponent, C chain precursor.  
 GN C1Qc OR C1Qc.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 CX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=Monocytes;  
 RX MEDLINE=91174759; PubMed=1706597;  
 RA Sellar G.C., Blake D.J., Reid K.B.M.;  
 RT "Characterization and organization of the gene encoding the A-, B-  
 RT and C-chains of human complement subcomponent C1q. The complete  
 RT derived amino acid sequence of human C1q.";  
 RT Biochem. J. 274:481-490 (1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Suzuki Y.,  
 RA Hata H., Nakagawa K., Mizuno S., Morinaga M., Kawamura M.,  
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A.,



RA Kawakami B., Nagai K., Isogai T., Sugano S.:  
 RA "NEBO human CDNA sequencing project."  
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Striplin M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
 RA Brownstein M.J., Uebin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Rana S.S., Loquellano N.A., Peters G.J., Adamson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Feley J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Boultard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalski U., Smallus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RA "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [4]  
 RP SEQUENCE OF 29-122.  
 RX MEDLINE=80020137; PubMed=486087;  
 RA Reid K.B.M.,  
 RT "Complete amino acid sequences of the three collagen-like regions  
 RT present in subcomponent C1q of the first component of human  
 RT complement."  
 RL Biochem. J. 179:367-371(1979).  
 RN [5]  
 RP REVIEW OF C1Q DEFICIENCY.  
 RX MEDLINE=98450587; PubMed=9777412;  
 RA Petry F.,  
 RT "Molecular basis of hereditary C1q deficiency."  
 RL Immunobiology 199:286-294(1998).  
 CC -1- FUNCTION: C1Q ASSOCIATES WITH THE PROENZYMES C1R AND C1S TO YIELD  
 CC C1, THE FIRST COMPONENT OF THE SERUM COMPLEMENT SYSTEM. THE  
 CC COLLAGEN-LIKE REGIONS OF C1Q INTERACT WITH THE CA(2+)-DEPENDENT  
 CC C1R(2)C1S(2) PROENZYME COMPLEX, AND EFFICIENT ACTIVATION OF C1  
 CC TAKES PLACE ON INTERACTION OF THE GLOBULAR HEADS OF C1Q WITH THE  
 CC FC REGIONS OF IGG OR IGM ANTIBODY PRESENT IN IMMUNE COMPLEXES.  
 CC -1- SUBUNIT: C1 IS A CALCIUM-DEPENDENT TRIMOLECULAR COMPLEX OF C1Q, R  
 CC AND S IN THE MOLAR RATION OF 1:2:2. C1Q SUBCOMPONENT IS COMPOSED  
 CC OF NINE SUBUNITS, SIX OF WHICH ARE DISULFIDE-LINKED DIMERS OF THE  
 CC A AND B CHAINS, AND THREE OF WHICH ARE DISULFIDE-LINKED DIMERS OF THE  
 CC C CHAIN.  
 CC -1- PM: O-linked glycans consist of Glc-Gal disaccharides bound to  
 CC the oxygen atom of post-translationally added hydroxyl groups.  
 CC -1- DISEASE: Defects in C1QG are a cause of C1q deficiency  
 CC [MIM:120575]. It is a rare genetic disorder which is associated  
 CC with recurrent infections and a high prevalence of lupus  
 CC erythematosus-like symptoms. It is characterized by a loss of  
 CC activation of the complement classical pathway.  
 CC -1- SIMILARITY: Contains 1 collagenous domain.  
 CC -1- SIMILARITY: Contains 1 C1Q domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 CC EMBL; AK057792; BAB71575.1; -  
 CC EMBL; BC009016; AA009016.1; -  
 CC PIR; S1451; C1HQC.

DR Genew; HGNC:1245; C1QG.  
 DR MIM; 120575; -  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003793; F:defense/immunity protein activity; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR001073; C1q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1q; 1.  
 DR Pfam; PF01391; Collagen; 1.  
 DR PRINTS; PR00007; COMPLEMENTC1Q.  
 DR SMART; SM00110; C1Q; 1.  
 DR PROSITE; PS01113; C1Q; 1.  
 KW Complement pathway; Plasma; Hydroxylation; Glycoprotein; Collagen;  
 KW Repeat; Signal; Disease mutation.  
 FT SIGNAL 1 28  
 FT CHAIN 29 245  
 FT DOMAIN 31 112  
 FT DOMAIN 113 245  
 FT DISULFID 32 36  
 FT MOD\_RES 39 39  
 FT MOD\_RES 42 42  
 FT MOD\_RES 45 45  
 FT MOD\_RES 54 54  
 FT MOD\_RES 57 57  
 FT MOD\_RES 63 63  
 FT MOD\_RES 66 66  
 FT MOD\_RES 71 71  
 FT MOD\_RES 75 75  
 FT CARBOHYD 75 75  
 FT MOD\_RES 81 81  
 FT MOD\_RES 84 84  
 FT CARBOHYD 84 84  
 FT MOD\_RES 93 93  
 FT MOD\_RES 96 96  
 FT MOD\_RES 99 99  
 FT MOD\_RES 105 105  
 FT MOD\_RES 105 105  
 FT VARIANT 43 43  
 FT CONFLICT 14 14 K -> R (IN REF. 2).  
 FT CONFLICT 23 23 P -> A (IN REF. 1).  
 FT CONFLICT 57 57 K -> P (IN REF. 4).  
 FT CONFLICT 66 66 P -> K (IN REF. 4).  
 FT CONFLICT 72 72 K -> P (IN REF. 4).  
 FT CONFLICT 84 84 P -> D (IN REF. 4).  
 FT CONFLICT 87 87 N -> K (IN REF. 4).  
 FT CONFLICT 90 90 M -> N (IN REF. 4).  
 FT CONFLICT 215 215 E -> G (IN REF. 2).  
 SQ SEQUENCE 245 AA; 25774 MW; FA17117EB7ABFC12 CRC64;  
 Query Match 19.0%; Score 170; DB 1; Length 245;  
 Best Local Similarity 26.6%; Pred. No. 7.1e-09;  
 Matches 46; Conservative 27; Mismatches 56; Indels 44; Gaps 6;  
 QY 9 GPFA-----PEP-----SFAFASTRSLVSGDAGPGRHQP-----L 42  
 DB 91 GPFGMPGVGPMPGIPPEPGEGRYKQKQSVFVTR-----QTHQPPARNSLI 138  
 QY 43 APFTBEVNTIGGDFDAAGVFRCLPGAYFFSFTLGLPKRTLSVKLMKRNDEVOAMITDD 102  
 DB 139 RFAVAVLNPGQDVTSTGKFTCKVPGIYYFVYHAS---HTANLCVLLVRSRGVAVTFCG 194  
 QY 103 GASRRREMOSQWMLARSGDAVWMLSHDDGAGVANSNGKXITTFSGFLVYPD 155  
 DB 195 HRSKTNQVNSGVLLRLQVGEVWLVANDI--YDMVGIGQSDSVFSGFLVFPD 245  
 RESULT 15  
 ID CA1A-CHICK STANDARD; PRT; 674 AA.  
 AC P08125;  
 DT 01-NOV-1988 (Rel. 08, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)

DT 15-JUL-1999 (Rel. 38, last annotation update)  
 DE Collagen alpha 1(X) chain precursor.  
 GN COL10A1.  
 OS Gallus gallus (chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OC NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE OF 48-674 FROM N.A. AND SEQUENCE OF 103-117 AND 453-466.  
 RX MEDLINE=86168227; PubMed=3082876;  
 RA Nishimura Y., Gordon M., van der Rest M., Schmid T., Linsemayer T.,  
 RA Olsen B.R.;  
 RT "The developmentally regulated type X collagen gene contains a long  
 RT open reading frame without introns";  
 RL J. Biol. Chem. 261:5041-5050(1986).  
 RN [2]  
 RP SEQUENCE OF 1-75 FROM N.A.  
 RX MEDLINE=89054019; PubMed=2461368;  
 RA Luvall P., Nishimura Y., Rosenblum N.D., Olsen B.R.;  
 RT "The type X collagen gene. Intron sequences split the 5'-untranslated  
 RT region and separate the coding regions for the non-collagenous amino-  
 RT terminal and triple-helical domains";  
 RL J. Biol. Chem. 263:18378-18385(1988).  
 RN [3]  
 RP REVISIONS TO C-TERMINUS.  
 RX MEDLINE=89380199; PubMed=2476437;  
 RA Yamaguchi N., Benay P.D., van der Rest M., Nishimura Y.;  
 RT "The cloning and sequencing of alpha 1(VIII) collagen cDNAs  
 RT demonstrate that type VIII collagen is a short chain collagen and  
 RT contains triple-helical and carboxyl-terminal non-triple-helical  
 RT domains similar to those of type X collagen.";  
 RL J. Biol. Chem. 264:16022-16029(1989).  
 CC -I- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTROPHIC  
 CC CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE  
 CC MINERALIZATION ZONES OF HYALINE CARTILAGE.  
 CC -I- SUBUNIT: Homotrimer.  
 CC -I- PTM: Prolines at the third position of the tripeptide repeating  
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -I- SIMILARITY: STRONG, TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.  
 CC -I- SIMILARITY: Contains 1 C1Q domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M13496; AAA48736.1; ALT\_SEQ.  
 DR EMBL: J04194; AAA48634.1; -.  
 DR PIR: S23297; S23297.  
 DR InterPro: IPR001073; C1Q.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF00386; C1q; 1.  
 DR Pfam: PF01391; Collagen; 8.  
 DR PRINTS: PR00007; COMPLEMENTC1Q.  
 DR ProDom: PD000007; C1q\_helix; 1.  
 DR SMART: SM00110; C1Q; 1.  
 DR PROSITE: PS01113; C1Q; 1.  
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Cartilage; Collagen; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 674  
 FT DOMAIN 19 52 NONHELICAL REGION (NC2).  
 FT DOMAIN 53 512 TRIPLE-HELICAL REGION.  
 FT DOMAIN 513 674 NONHELICAL REGION (NC1).  
 FT DOMAIN 539 674 C1Q.  
 FT MOD\_RES 453 453 HYDROXYLATION.  
 FT MOD\_RES 456 456 HYDROXYLATION.  
 FT SEQUENCE 674 AA; 66434 MW; EAB48B1EF17AB145 CRC64;

Query Match 18.6%; Score 166.5; DB 1; Length 674;  
 Best local Similarity 28.9%; Pred. No. 5e-08;  
 Matches 50; Conservative 19; Mismatches 71; Indels 33; Gaps 5;  
 QY 9 GPAPPEP-----RFAFAARTRSIVG-----SDAGGPR 38  
 DB 501 GPQPPPEPQOSTIPGCVKGESRELSGMSFMKAGANQALTGPVSAFTYILSKAYIGAT 560  
 QY 39 HOPLAEPTEFVNIGDPDAAGVFRCLPGAYFEFSTGLPKRTLSVKLMKRRDEVQAM 98  
 DB 561 -VPIKFPKILYNQOHDPRTGIFTCRIPGLYFYSYHV-HAKGTNVVALYKQSPVM-Y 617  
 QY 99 IYDQASRRRQSQSVMLALRRGDVWLSHHDGCGAYSNHGYTTESGFL 151  
 DB 618 TYDEYQKGYLDQASGSAVIDLMENDQWVLQLPNSSENSGLYSSEYVHSSFSGL 670

Search completed: January 12, 2004, 08:12:59  
 Job time : 7.10778 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:10 ; Search time 23.4132 Seconds  
(without alignments)  
1873.686 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_160\_358  
Sequence: 1 ADADAPARGPAPPEPRSAF.....LYTPDLAPAPPGLAGSELL 170

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues  
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	894	100.0	329	4 Q81V25	Q81V25 homo sapien
2	807.5	90.3	205	11 Q9D0W2	Q9D0W2 mus musculu
3	807.5	90.3	205	11 Q8R066	Q8R066 mus musculu
4	801.5	89.7	205	11 Q9DCB6	Q9DCB6 mus musculu
5	230	25.7	158	4 Q9H667	Q9H667 mus musculu
6	230	25.7	1077	4 Q8E711	Q8E711 homo sapien
7	226	25.3	158	11 Q8K110	Q8K110 mus musculu
8	220.5	24.7	182	11 Q8R1P2	Q8R1P2 mus musculu
9	220.5	24.7	281	11 Q9QXP7	Q9QXP7 mus musculu
10	208.5	23.3	243	4 Q8N6P2	Q8N6P2 mus musculu
11	197.5	22.1	243	11 Q8R002	Q8R002 mus musculu
12	194.5	21.8	243	11 Q8K479	Q8K479 mus musculu
13	189	21.1	247	11 Q8BRW2	Q8BRW2 mus musculu
14	187.5	21.0	294	11 Q9D8U4	Q9D8U4 mus musculu
15	185	20.7	240	6 Q95W04	Q95W04 bos taurus
16	180	20.1	244	11 Q8K3R4	Q8K3R4 rattus norv

17	179.5	20.1	120	11 Q8R1Z2	Q8R1Z2 mus musculu
18	179.5	20.1	264	11 Q8BKRO	Q8BKRO mus musculu
19	177	19.8	287	11 Q8CFR0	Q8CFR0 mus musculu
20	176	19.7	243	6 Q95J07	Q95J07 macaca mula
21	166	18.6	675	6 Q9N178	Q9N178 sus scrofa
22	163	18.2	224	4 Q8TUK8	Q8TUK8 homo sapien
23	163	18.2	224	4 Q8TUK8	Q8TUK8 homo sapien
24	161	18.0	224	11 Q8BGU2	Q8BGU2 mus musculu
25	161	18.0	246	6 Q62789	Q62789 sus scrofa
26	160.5	18.0	295	11 Q9ES30	Q9ES30 mus musculu
27	157.5	17.6	244	11 Q9Z1K4	Q9Z1K4 rattus norv
28	157.5	17.6	744	11 Q9Z1S8	Q9Z1S8 mus musculu
29	152	17.0	197	11 Q8HGO0	Q8HGO0 mus musculu
30	152	17.0	312	11 Q8CHX9	Q8CHX9 canis fami
31	151	16.9	194	6 Q95J95	Q95J95 mus musculu
32	148	16.6	196	11 Q9Z0N0	Q9Z0N0 tamias sibi
33	147.5	16.5	333	4 Q8TUD4	Q8TUD4 homo sapien
34	143	16.0	289	11 Q8BYD7	Q8BYD7 mus musculu
35	142	15.9	705	4 Q8TEJ5	Q8TEJ5 mus musculu
36	134	15.0	245	11 Q9DCM6	Q9DCM6 mus musculu
37	134	15.0	1017	11 Q99K41	Q99K41 mus musculu
38	133.5	14.9	195	11 Q8B2S3	Q8B2S3 mus musculu
39	130	14.5	164	11 Q64144	Q64144 rattus sp.
40	129	14.4	347	4 Q961H6	Q961H6 homo sapien
41	129	14.4	583	4 Q96G58	Q96G58 homo sapien
42	129	14.4	992	4 Q9UG76	Q9UG76 homo sapien
43	129	14.4	1016	4 Q9Y6C2	Q9Y6C2 homo sapien
44	127.5	14.3	198	11 Q8BMF0	Q8BMF0 mus musculu
45	126.5	14.1	198	11 Q8BME9	Q8BME9 mus musculu

ALIGNMENTS

RESULT 1  
ID Q81V25 PRELIMINARY; PRT; 329 AA.  
AC Q81V25;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DE Similar to C1q and tumor necrosis factor related protein 4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Strausberg R.;  
DR EMBL; BC035628; AAH35628.1; -  
SQ SEQUENCE 329 AA; 35256 MW; 16064DA8182A6732 CRC64;

Query Match 100.0%; Score 894; DB 4; Length 329;  
Best Local Similarity 100.0%; Pred. No. 2.9e-82;  
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ADADAPARGPAPPEPRSAFSAARTRSLVGSAGPGRPHOPLAFDTFVNIGDPPAAG	60
DB	160	ADADAPARGPAPPEPRSAFSAARTRSLVGSAGPGRPHOPLAFDTFVNIGDPPAAG	219
QY	61	VRRCRLPGAYFSPFTLGLKPRKTLGVKMKNDVQAMTYDDGASRRRMOQSVLALR	120
DB	220	VRRCRLPGAYFSPFTLGLKPRKTLGVKMKNDVQAMTYDDGASRRRMOQSVLALR	279
QY	121	RSDAVWLSHDHDGYGAYSNHGKYYTFSGFLVYPPDLAPAPPGLAGSELL	170
DB	280	RSDAVWLSHDHDGYGAYSNHGKYYTFSGFLVYPPDLAPAPPGLAGSELL	329
RESULT 2	Q9D0W2		

Query	Subject	Score	Length	Matches	Mismatches	Indels	Gaps
1	ADADAPARGPAPPEPRSAFSAARTSLVSGDAGPGRHOPLAFDTFVNIGGFDPAAG	90.3%	326	157	1	1	1
159	ADADAPARG-PAAPPEPRSAFSAARTSLVSGDAPGRHPLAFDTFVNIGGFDPAAG	92.9%	326	157	1	1	1
61	VFCRRLPGAYPFSTLTGLKTRKTLSTVLKMKRDPVQAMITDGGSRREMOSSQSVMLALR	807.5	120	120	0	0	0
218	VFCRRLPGAYPFSTLTGLKTRKTLSTVLKMKRDPVQAMITDGGSRREMOSSQSVMLALR	807.5	120	120	0	0	0
121	RGDAVWMLSHDHDGAYSNHGKITYTSGFLVYEDLPAAPPGGASBL	169	169	169	0	0	0
278	RGDAVWMLSHDHDGAYSNHGKITYTSGFLVYEDLPAAPPGGASBL	169	169	169	0	0	0

NCBI TaxID=10090;  
SEQUENCE FROM N.A.  
Strausberg R.;  
Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.  
EMBL: BC027315; AA427315.1;  
MGD: MGI:1914695; 0710001E10R1k.  
InterPro: IPR001073; Clq.  
Pfam: PF00386; Clq; 2.  
SMART: SM00110; Clq; 2.  
PROSITE: PS01113; Clq; 2.  
SEQUENCE 326 AA; 35057 MW; 72339172B7B1051A CRC64;

Query Match 90.3%; Score 807.5; DB 11; Length 326;  
Best Local Similarity 92.9%; Pred. No. 1,7e-73;  
Matches 157; Conservative 1; Mismatches 10; Indels 1; Gaps 1;

RESULT 4  
09DCB6 PRELIMINARY; PRT: 205 AA.  
ID 09DCB6  
AC 09DCB6  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE 0710001E10R1k protein.  
GN 0710001E10R1K.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=C57BL/6J; TISSUE=Brain.  
MEDLINE=21085660; PubMed=11217851;  
RA Kawai Y., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
RA Aizawa K., Iwata M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,  
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nakai I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Stubbli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Botelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustingich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima Y., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wyszewski Y., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.;  
[2]  
Functional annotation of a full-length mouse cDNA collection";  
RL Nature 409:685-690(2001).  
EMBL: AK002948; BAB22473.1;  
MGD: MGI:1914695; 0710001E10R1k.  
InterPro: IPR001073; Clq.  
Pfam: PF00386; Clq; 1.  
DR PRINTS: PR00007; COMPLEMENTC1Q.  
DR SMART: SM00110; Clq; 1.

DR PROSITE; PS01113; ClQ; 1.  
SQ SEQUENCE 205 AA; 22215 MM; 56AD37793C437300 CRC64;  
Query Match 89.7%; Score 801.5; DB 11; Length 205;  
Best Local Similarity 92.3%; Pred. No. 3.7e-73;  
Matches 156; Conservative 1; Mismatches 11; Indels 1; Gaps 1;  
QY 1 ADAAPARGPAPPEPRSAFSAARTSLVSGSDAGPGRHOPLADTFEYVNIIG 60  
DB 38 ADAAPARGP-PAAEPFRSAFSAARTSLVSGSDAGPGRHOPLADTFEYVNIIGSDPDAAG 96  
QY 61 VFCRLGAYFFSFTLGKLPKRTLSVKLMKRDDEVQAMITDDGASRRREMOSQSVMLALR 120  
DB 97 VFCRLGAYFFSFTLGKLPKRTLSVKLMKRDDEVQAMITDDGASRRREMOSQSVMLALR 156  
QY 121 RGDVWLLSHDHDGYGASNHGKXITTSGLVYFDLPAAPPGCASL 169  
DB 157 RGDVWLLSHDHDGYGASNHGKXITTSGLVYFDLPAAPPGCASL 205  
RESULT 5  
Q9H667 PRELIMINARY; PRT; 158 AA.  
ID Q9H667  
AC Q9H667  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Homo sapiens (Human).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Small intestine;  
RA Matanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,  
RA Suzuki Y., Ohayashi M., Nishi T., Shibahara T., Tanaka T.,  
RT "NEDO human cDNA sequencing project."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Uterus;  
RA Straubeberg R.;  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK026222; BAB15398.1; -  
DR EMBL; BC007520; AA07520.1; -  
DR InterPro: IPR001073; ClQ.  
DR Pfam; PF00386; ClQ; 1.  
DR PRINTS; PR00007; COMPLEMENTC1Q.  
DR SMART; SM00110; ClQ; 1.  
DR PROSITE; PS01113; ClQ; 1.  
KM Hypothetical protein.  
SQ SEQUENCE 158 AA; 17625 MM; 47DB10EDD6DC9760 CRC64;  
Query Match 25.7%; Score 230; DB 4; Length 158;  
Best Local Similarity 38.4%; Pred. No. 2.1e-15;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;  
QY 4 DAPARGP-----PAPPEPRSAFSAARTSLVSGSDAGPGRHOPLADTFEYVNIIG 53  
DB 5 DAPVTPAATILPVHYVPLPQMKRVAFSAART-----SNLAPGTLDDPIVFDLLNNLGE 59  
QY 54 DFDAAAGVFCRLGAYFFSFTLGKLP-RKTLISVKLMKRDDEVQAMITDDGASRRREMOS 112  
DB 60 TFDLQGRFNCVPNGVYVFIHMLKLVANNVPLVNLKNEBEVLVSAVANDGAP-DHETAS 118  
QY 113 QSVMLALRGDVAWLLSHDHDGYGA-YSNHGKXITTSGLVYFD 155  
DB 119 NHAILQFGQDQIWLRLH---RGAIVGSSWKYSTFGYLLYOD 158  
RESULT 6

Q8TE71 PRELIMINARY; PRT; 1077 AA.  
ID Q8TE71  
AC Q8TE71  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE EBI1.  
GN EBI1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Aarabainat W., Miller J.L.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY074490; AAL71549.1; -  
DR InterPro: IPR001073; ClQ.  
DR Pfam; PF00386; ClQ; 1.  
DR SMART; SM00110; ClQ; 1.  
DR PROSITE; PS01113; ClQ; 1.  
SQ SEQUENCE 1077 AA; 120974 MM; 2B88BF3C47D032D6 CRC64;  
Query Match 25.7%; Score 230; DB 4; Length 1077;  
Best Local Similarity 38.4%; Pred. No. 2.5e-14;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;  
QY 4 DAPARGP-----PAPPEPRSAFSAARTSLVSGSDAGPGRHOPLADTFEYVNIIG 53  
DB 924 DAPVTPAATILPVHYVPLPQMKRVAFSAART-----SNLAPGTLDDPIVFDLLNNLGE 978  
QY 54 DFDAAAGVFCRLGAYFFSFTLGKLP-RKTLISVKLMKRDDEVQAMITDDGASRRREMOS 112  
DB 979 TFDLQGRFNCVPNGVYVFIHMLKLVANNVPLVNLKNEBEVLVSAVANDGAP-DHETAS 1037  
QY 113 QSVMLALRGDVAWLLSHDHDGYGA-YSNHGKXITTSGLVYFD 155  
DB 1038 NHAILQFGQDQIWLRLH---RGAIVGSSWKYSTFGYLLYOD 1077  
RESULT 7  
Q8KI10 PRELIMINARY; PRT; 158 AA.  
ID Q8KI10  
AC Q8KI10  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Similar to hypothetical protein FLJ22569.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Uterus;  
RA Straubeberg R.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC027523; AA07523.1; -  
DR InterPro: IPR001073; ClQ.  
DR Pfam; PF00386; ClQ; 1.  
DR PRINTS; PR00007; COMPLEMENTC1Q.  
DR SMART; SM00110; ClQ; 1.  
DR PROSITE; PS01113; ClQ; 1.  
KM Hypothetical protein.  
SQ SEQUENCE 158 AA; 17533 MM; 86E9321C9225FCB CRC64;  
Query Match 25.3%; Score 226; DB 11; Length 158;  
Best Local Similarity 37.8%; Pred. No. 5.3e-15;  
Matches 62; Conservative 17; Mismatches 63; Indels 22; Gaps 6;  
QY 4 DAPARGP-----PAPPEPRSAFSAARTSLVSGSDAGPGRHOPLADTFEYVNIIG 53  
DB 5 DAPVTPAATILPVHYVPLPQMKRVAFSAART-----SNLAPGTLDDPIVFDLLNNLGE 59

QY 54 DFDAAAGVFCRCLPGAYFFSFTLGLP-RKTLVSKLMKNEDEVQAMTYDDGASRRRMOQ 112  
 DB 60 TFDNQLGRFCNCPVNGTVTFTHMLKLVNVPYVNMKNKEEVLSAANAGAP-DHETAS 118  
 QY 113 QSVMLALRGDAVWLLSHDHGCGA-YSNHGKTYTFSGFLVYPD 155  
 DB 119 NHAVALQLQGGQVLRH---RCATGSSWKYSTFGYLLYQD 158

RESULT 8  
 Q8R1P2 PRELIMINARY; PRT; 182 AA.  
 AC Q8R1P2;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Similar to RIKEN cDNA 1600017K21 gene (fragment).  
 GN 1600017K21RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RL Submitted (Feb-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC023468; AAH23468.1;  
 DR MGD; MGI:1919254; 1600017K21RIK.  
 DR InterPro; IPR001073; C1q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1q; 1.  
 DR SMART; SMO0110; C1Q; 1.  
 FT NON\_TER  
 SQ SEQUENCE 182 AA; 20863 MW; 559C73DE9517882F CRC64;

Query Match 24.7%; Score 220.5; DB 11; Length 182;  
 Best Local Similarity 39.2%; Pred. No. 2,3e-14;  
 Matches 56; Conservative 20; Mismatches 58; Indels 9; Gaps 4;

QY 18 SAFSAARTSLVSGDAGPRHOPLPDTEFVNIGDGDAAAGVFCRLPGAYFFSFTLG 77  
 DB 47 AAFSVGRKKALHNSD-----YFQPVVDTFENLYKHFNMFTGKFCYVCGIYFSSLVNH 101  
 QY 78 KLPRKTLVSKLMKNEDEVQAMTYDDGASRRRMOQSVMLALRGDAVW--LLSHDHGY 135  
 DB 102 TWNOKETVLMKNKEEV-VILVQ-VSDRSIMOSQSLMELREDEVWVRLFKGERENA 159  
 QY 136 GAVSNHGKTYTFSGFLVYPDLP 158  
 DB 160 IFSDEPDYITTFSGYLVKPASEP 182

RESULT 9  
 Q9QXP7 PRELIMINARY; PRT; 281 AA.  
 AC Q9QXP7;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative secreted protein ZS1G37 (1600017K21RIK protein).  
 GN ZS1G37 OR 1600017K21RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RL Submitted (May-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC029485; AAH29485.1;  
 DR InterPro; IPR001073; C1q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1q; 1.  
 FT Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.

RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Placenta;  
 RX MEDLINE=21085660; Pubmed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Atakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schmitt L.M., Staudt P., Suzuki R., Tomita M., Wagner L., Wachi T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Bash G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Monbaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,  
 RA Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,  
 RA Hayaishizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL; AF192499; AAF06664.1;  
 DR EMBL; AK005484; BAB24070.1;  
 DR MGD; MGI:1919254; 1600017K21RIK.  
 DR InterPro; IPR001073; C1q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1q; 1.  
 DR Pfam; PF01391; Collagen; 1.  
 DR PRINTS; PR00007; COMPLEMENTC1Q.  
 DR SMART; SMO0110; C1Q; 1.  
 SQ SEQUENCE 281 AA; 32009 MW; C9816216DB6419E2 CRC64;

Query Match 24.7%; Score 220.5; DB 11; Length 281;  
 Best Local Similarity 39.2%; Pred. No. 4e-14;  
 Matches 56; Conservative 20; Mismatches 58; Indels 9; Gaps 4;

QY 18 SAFSAARTSLVSGDAGPRHOPLPDTEFVNIGDGDAAAGVFCRLPGAYFFSFTLG 77  
 DB 146 AAFSVGRKKALHNSD-----YFQPVVDTFENLYKHFNMFTGKFCYVCGIYFSSLVNH 200  
 QY 78 KLPRKTLVSKLMKNEDEVQAMTYDDGASRRRMOQSVMLALRGDAVW--LLSHDHGY 135  
 DB 201 TWNOKETVLMKNKEEV-VILVQ-VSDRSIMOSQSLMELREDEVWVRLFKGERENA 258  
 QY 136 GAVSNHGKTYTFSGFLVYPDLP 158  
 DB 259 IFSDEPDYITTFSGYLVKPASEP 281

RESULT 10  
 Q8N6P2 PRELIMINARY; PRT; 243 AA.  
 AC Q8N6P2;  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE C1q and tumor necrosis factor related protein 5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RL Submitted (May-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC029485; AAH29485.1;  
 DR InterPro; IPR001073; C1q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1q; 1.

DR Pfam: PF01391; Collagen; 1.  
DR PRINTS: PR00007; COMPLEMENTC1Q.  
DR SMART: SM00110; C1Q; 1.  
SQ SEQUENCE 243 AA; 25326 MW; 6D9306A0EB21B44A CRC64;

Query Match 23.3%; Score 208.5; DB 4; length 243;  
Best Local Similarity 35.7%; Pred. No. 5; Ae 13;  
Matches 56; Conservative 18; Mismatches 72; Indels 11; Gaps 3

QY 4 DAPARGPAPP-----EPRSAFSAARTSLVSGDAGPGPRHQPPLAFDTFEVNIIGDFDAA 58  
DB 85 EAGPAGPFGPAGGECVPPRPSAFSAKRSSRV-----PPSPDAFLPFDRLVLNVEGHYDAV 139

QY 59 AGVPRCLPRGAYFSPFTLGKLPKRLTSLVYLMKNRBEVQAMTYDDCASRRKMQSQSVMLA 118  
DB 140 TGETTCQVGVGYFFA-VHATYVRAFLQPLDVNLNGESIASFQFQGWGPKPASLSCGAMVR 198

QY 119 LRRGDVWLISHDHQVGAYSNHGKYLIFSGFLVYPD 155  
DB 199 LEPEDQVWVQVGVGDYITGIYASIKTDSIFSGFLVYSD 235

Query Match	Best local Similarity	Matches	Score	DB	Length
22.1%	33.3%	16	197.5	DB 11	243
Conservative		71			
Mismatches		71			
Indels		23			
Gaps		3			

```

ID      Q8K479.          PRELIMINARY;          PRT;          243 AA.
AC      Q8K479.
DT      01-OCT-2002 (TREMBLrel. 22, Created)
DT      01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT      01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE      Complement-c1q tumor necrosis factor-related protein.
GN      C1QTNF5.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxId=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C57BL/6J;
RX      MEDLINE=22135657; PubMed=12140190;
RA      Kaneva S., Hawes N.L., Chang B., Heckenlively J.R., Naggert J.K.,
RA      Nishina P.M.;
RT      "Mfcp, a gene encoding a frizzled related protein, is mutated in the
RT      mouse retinal degeneration 6.";
RL      Hum. Mol. Genet. 11:1879-1886(2002).
DR      EMBL; AF469650; AAM89217.1; -.
DR      MGD; MGI:2385958; C1qtnf5.
DR      InterPro; IPR001073; C1q.
DR      InterPro; IPR000087; Collagen.
DR      Pfam; PF00386; C1q; 1.
DR      Pfam; PF01391; Collagen; 2.
DR      PRINTS; PRO0007; COMPLEMENTC1Q.
DR      SMART; SMO0110; C1Q; 1.
SQ      SEQUENCE 243 AA; 25436 MW; 9FAD5804349791D9 CRC64;

```

Query Match	21.8%;	Score 194.5;	DB 11;	Length 243;
Best Local Similarity	32.7%;	Pred. NO. 1.4e-11;		
Matches 54;	Conservative 17;	Mismatches 71;	Indels 23;	Gaps 3

```

QY      8  RGPAP-----DEPRASFAARTSLVGSADAGGPHOPAFPTFEFN  50
Db      77  KGEQGPGEAGPMGALGPGAGCSVPNRSASFAKSESKV-----PPMDTLPFPRVLIN  131

QY      51  IGDGPDAAAGVFCRLPGAYPFPSFTLGLPKRTLSVKMKNRDEQAMTYDDGASRRREM  110
Db      132  EQGHYDPTTGTCQVPGVYTFYA-VHATVYASLSLOFLVKAGQSTIASFFQYFGGMPKPPAS  190

QY      111  QSQSVMTLALRRGDAWLLSHDHGDSGAYASNHGKYYITFSGLFVYPD  155
Db      191  LSGAMVRLPEEDOVWQVGVGDYITGIYASIKTSTFSGFLVYSD  225

```

RESULT	13
Q8BRW2	
ID	Q8BRW2
AC	Q8BRW2
DT	01-MAR-2003 (TREMBlrel. 23, Created)
DT	01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT	01-MAR-2003 (TREMBlrel. 23, last annotation update)
DE	Adipocyte complement related protein of 30 kDa.
OS	Mus musculus (Mouse).
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxId=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RX	MEDLINE=22354683; PubMed=12466851;
RA	The FANTOM Consortium,
RA	The RIKEN Genome Exploration Research Group Phase I & II Team;
RT	"Analysis of the mouse transcriptome based on functional annotation of
RL	60,770 full-length cDNAs."
RL	Nature 420:553-573(2002).
EMBL	AK041214; BAC00866.1; -
DR	SEQUENCE 247 AA; 26751 MW; 0D3F64C789CAEF3 CRC64;
SO	SEQUENCE

RESULT 12  
Q8K479

Query Match	21.1%;	Score 189;	DB 11;	Length 247;
Best Local Similarity	33.5%;	Pred. No. 5.3e-11;		

Matches 53; Conservative 24; Mismatches 73; Indels 8; Gaps 5;

QY 3 ADAPARGPAPP-----EPKSAFSAARTRSLVSGDAGCPGRHQPLAFDTEFVNIQGDPA 58  
 DB 91 AEGP-RGFPPTGPKRGEGEREAAYVRSFVGLERVTVPVPIRFTKIFVNOQNHVGS 149  
 QY 59 AGVFRCLPGAYFFSFTLGKLPKRTLSVKLMKNRDEVOAMTYDGCASRRRMOSSVWLA 118  
 DB 150 TGEFYCNIPGLYFESYHI-TYMKDKVKSLFK-KDKAVLFTYDQYQEKVNDQASGVLLH 207

QY 119 LRGGDAVWLSSH-DHGDGAYSNHCKYITPFGFLVYPD 155  
 DB 208 LEVGDQVWLQYVGEENNGVADVNDSTFTGFLVYLD 245

RESULT 14  
 Q9DBU4 PRELIMINARY; PRT; 294 AA.  
 AC Q9DBU4  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE 1810033K05RIK protein (RIKEN CDNA 1810033K05 gene).  
 GN 1810033K05RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Pancreas;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,  
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Glass C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schmitt L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Guetincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyshak-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlschki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection";  
 RL Nature 409:685-690(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC Straube R.;  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK007683; BAB25187.1; -  
 DR EMBL; BC030324; AAH30324.1; -  
 DR MGP; MGI:1916433; 1810033K05RIK.  
 DR InterPro; IPR001073; Clq.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; Clq; 1.  
 DR Pfam; PF01391; Collagen; 2.  
 DR PRINTS; PR00007; COMPLEMENTC1Q.  
 DR PROSITE; PS01113; Clq; 1.  
 SQ SEQUENCE 294 AA; 30865 MW; 6D3905AE7C19E6FA CRC64;

Query Match 21.0%; Score 187.5; DB 11; Length 294;  
 Best Local Similarity 32.5%; Pred. No. 9.3e-11;  
 Matches 52; Conservative 30; Mismatches 53; Indels 25; Gaps 8;

QY 8 RGPAPPP-----RSAPSAARTRSLVSGDAGCPGRHQ-PLAFDTEFVNIQGDPA 58  
 DB 141 KGEPLGPGSCGSSRAKSAFSAVATKSY-----PRRLPIKPKILMEGSHYNAS 192  
 QY 59 AGVFRCLPGAYFFSFTLGKLPKRTLSVKLMKNRDEVOAMTYDGCASRRRMOSSVWLA 118  
 DB 193 SGKRVCSVPGLYFTYDI-FLANKHLAIGLVH-GQYRIRTPDANTG-NHDVAGSTILA 249  
 QY 119 LRGGDAVWLSSH-DHGDGAYSNHCKYITPFGFLVYPD 155  
 DB 250 LKEGDVWLQI FYSEQNGLF--YDPWTDSLFTGFLVYAD 287

RESULT 15  
 Q95MQ4 PRELIMINARY; PRT; 240 AA.  
 AC Q95MQ4  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Adipose tissue-specific protein adipo Q.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 NX NCBI\_Taxid=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=21369933; PubMed=11382781;  
 RX Sato C., Yasukawa Z., Honda N., Matsuda T., Kitajima K.;  
 RA "Identification and Adipocyte Differentiation-dependent Expression of  
 RT the Unique Distal Acid Residue in an Adipose Tissue-specific  
 RT Glycoprotein, Adipo Q.";  
 RL J. Biol. Chem. 276:28849-28856(2001).  
 DR EMBL; AF269230; AAK58902.1; -  
 DR InterPro; IPR001073; Clq.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; Clq; 1.  
 DR Pfam; PF01391; Collagen; 1.  
 DR PRINTS; PR00007; COMPLEMENTC1Q.  
 DR PRODOM; PD000007; Collagen; 1.  
 DR SMART; SM00110; Clq; 1.  
 DR PROSITE; PS01113; Clq; 1.  
 KW Collagen.  
 SQ SEQUENCE 240 AA; 26091 MW; C6253BA803B9A668 CRC64;

Query Match 20.7%; Score 185; DB 6; Length 240;  
 Best Local Similarity 31.4%; Pred. No. 1.3e-10;  
 Matches 50; Conservative 29; Mismatches 72; Indels 8; Gaps 5;

QY 3 ADAPARGPAPP-----EPKSAFSAARTRSLVSGDAGCPGRHQPLAFDTEFVNIQGDPA 58  
 DB 83 AEGP-RGFPPTGPKRGEGEREAAYVRSFVGLERVTVPVPIRFTKIFVNOQNHVGS 141  
 QY 59 AGVFRCLPGAYFFSFTLGKLPKRTLSVKLMKNRDEVOAMTYDGCASRRRMOSSVWLA 118  
 DB 142 TGEFYCNIPGLYFESYHI-TYMKDKVKSLFK-KDKAVLFTYDQYQEKVNDQASGVLLH 199  
 QY 119 LRGGDAVWLSSH-DHGDGAYSNHCKYITPFGFLVYPD 156  
 DB 200 LEVGDQVWLQYVGEENNGVADVNDSTFTGFLVYINI 238

Search completed: January 12, 2004, 08:19:09  
 Job time : 23.4132 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:09 ; Search time 25.6188 Seconds

(without alignment)  
1053.272 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_160\_358

Perfect score: 894  
Sequence: 1 ADADAPARGPPAPPPPSAF.....LVYDLPAPAPPGIGASELL 170

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

\_A\_Geneseq\_19Jun03.\*  
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	894	100.0	329	22 AAB61606 Human ZACRP4. Hom
2	894	100.0	329	23 AAB61606 Human novel secret
3	871	97.4	329	22 AAB61424 Monkey MANGO 245 p
4	743	83.1	192	22 AAB61479 Human MANGO 245 cy
5	743	83.1	334	22 AAB61466 Human MANGO 245 ma
6	743	83.1	348	22 AAB61423 Human MANGO 245 pr
7	730	81.7	299	23 AAB61385 Adipocyte compleme
8	696	77.9	221	23 AAB61385 Adipocyte compleme
9	683	76.4	134	23 AAB61473 Monkey MANGO 245 C

10	666.5	74.6	284	23 AAB61384 Adipocyte compleme
11	664	74.3	127	22 AAB61578 Human secreted pro
12	609	68.1	117	22 AAB61470 Human MANGO 245 C1
13	482	53.9	130	22 AAB61468 Mature monkey MANG
14	432	48.3	133	23 AAB61386 Adipocyte compleme
15	357.5	40.0	109	22 AAB61488 Murine MANGO 245 p
16	321	35.9	126	22 AAB61472 Murine MANGO 245 C
17	320	35.8	125	22 AAB61477 Human MANGO 245 ex
18	320	35.8	126	22 AAB61469 Human MANGO 245 C1
19	318	35.6	126	22 AAB61489 Murine MANGO 245 C
20	305.5	34.2	123	21 AAB62189 Human OREF ORF1953
21	237.5	26.6	252	22 AAB69590 Human zacrif5. Hom
22	237.5	26.6	252	22 AAO19412 Human molecule for
23	237.5	26.6	252	23 AAO83100 Human secreted pro
24	237.5	26.6	252	23 AAB53292 Human polypeptide
25	230	25.7	202	22 AAM9927 Human polypeptide
26	230	25.7	709	23 AAB08645 Human pancreatic c
27	230	25.7	710	23 AAO15420 Human genest metab
28	230	25.7	746	20 AA129512 Human lung tumour
29	230	25.7	746	21 AAB4461 Human lung tumour
30	230	25.7	746	22 AAE13802 Human lung tumour
31	230	25.7	800	22 AAM78334 Human protein SEO
32	230	25.7	908	22 AAE09839 Novel human protei
33	230	25.7	909	22 AAE09845 Novel human protei
34	230	25.7	957	22 AAE09841 Novel human protei
35	230	25.7	958	22 AAE09847 Novel human protei
36	230	25.7	992	22 AAE09843 Novel human protei
37	230	25.7	993	22 AAE09849 Novel human protei
38	230	25.7	1042	22 AAE09848 Novel human protei
39	230	25.7	1043	22 AAE09844 Novel human protei
40	230	25.7	1091	22 AAE09840 Novel human protei
41	230	25.7	1092	22 AAE09846 Novel human protei
42	230	25.7	1136	22 AAE09842 Novel human protei
43	230	25.7	1137	22 AAE09848 Novel human protei
44	230.5	24.7	281	20 AAW98014 Mouse adipocyte-SP
45	220.5	24.7	281	21 AAB08428 Amino acid sequenc

## ALIGNMENTS

RESULT 1	
AAB61606	AAB61606 standard; Protein; 329 AA.
ID	
XX	
AC	AAB61606;
XX	
DT	05-APR-2001 (first entry)
XX	
DE	Human ZACRP4.
XX	
KW	Human; zacrif4; complement factor C1q domain; chromosome 11q11;
KW	energy balance; cellular metabolic reaction; autocrine factor;
KW	development; cell proliferation; differentiation; cell survival.
XX	
OS	Homo sapiens.
XX	
FT	Key
FT	Peptide
FT	Domain
FT	Domain
FT	Domain
FT	Domain
XX	
XX	WO200102565-A2.
XX	
XX	11-JAN-2001.
XX	
XX	28-JUN-2000; 2000MO-US17692.
XX	
XX	01-JUL-1999; 99US-0346502.
XX	

PA (ZYMO) ZYMOGENETICS INC.  
 XX  
 PI Holloway JL, Lok S;  
 XX WPI, 2001-138140/14.  
 DR N-PSDB; AAF28672.  
 XX  
 PT Novel secreted protein ZACRP4 polypeptides having tandem C1q globular  
 PT domains, useful for studying cell-cell communication and regulation of  
 PT cellular processes -  
 XX  
 PS Claim 1; Page 77-78; 82pp; English.  
 XX  
 CC The present sequence is human ZACRP4 protein. ZACRP4 protein has two  
 CC complement factor C1q domains. The ZACRP4 gene is located on human  
 CC chromosome 11q11. The ZACRP4 coding sequence and protein have a number of  
 CC uses described in the specification, including, modulation of energy  
 CC balance and cellular metabolic reactions in mammals. In addition, ZACRP4  
 CC protein is useful as an autocrine factor, particularly during  
 CC development, in mediating the processes of an organism, in regulating  
 CC cellular processes such as cell proliferation and/or differentiation,  
 CC cell survival and energy balance.  
 CC  
 XX  
 SQ Sequence 329 AA;  
 Query Match 100.0%; Score 894; DB 22; Length 329;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-89;  
 Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ADADAPARGPAPPEPRSAFSAARTSLVSGDAGPGRHOPLAEDTEFVNIGDPDAAAG 60  
 DB 160 ADADAPARGPAPPEPRSAFSAARTSLVSGDAGPGRHOPLAEDTEFVNIGDPDAAAG 219  
 QY 61 VFRCRLPGAYFFSFTLGLPKRTLSVKLMKRDVQAMITYDDGASRRREMOSQSVMLALR 120  
 DB 220 VFRCRLPGAYFFSFTLGLPKRTLSVKLMKRDVQAMITYDDGASRRREMOSQSVMLALR 279  
 QY 121 RGDVWMLSHDHGAGYASNHGKTYTFSGFLVYDPLAPAPPGIGASELL 170  
 DB 280 RGDVWMLSHDHGAGYASNHGKTYTFSGFLVYDPLAPAPPGIGASELL 329

RESULT 2  
 ABG79643  
 ID ABG79643 standard; Protein; 329 AA.

AC ABG79643;

DT 15-NOV-2002 (first entry)

DE Human novel secreted protein SECP19, Incyte ID No. 931619CD1.

XX Human; SECP; secreted protein; cell proliferative disorder;  
 KW actinic keratosis; arteriosclerosis; bursitis; hepatitis; cancer;  
 KW autoimmune disorder; inflammatory disorder; AIDS; asthma; allergy;  
 KW acquired immunodeficiency syndrome; anaemia; atopic dermatitis;  
 KW cardiovascular disorder; congestive heart failure; vascular tumour;  
 KW ischaemic heart disease; myocardial infarction; epilepsy; stroke;  
 KW hypertensive heart disease; neurological disorder; cerebral neoplasm;  
 KW Alzheimer's disease; developmental disorder; renal tubular acidosis;  
 KW Cushing's syndrome; Duchenne muscular dystrophy; hypothyroidism;  
 KW Becker muscular dystrophy.

OS Homo sapiens.

FN WO200262841-A2.

PD 15-AUG-2002.

PF 28-JAN-2002; 2002WO-US02616.

XX 02-FEB-2001; 2001US-266195P.

PR 08-FEB-2001; 2001US-267924P.

PR 09-FEB-2001; 2001US-267816P.  
 PR 09-FEB-2001; 2001US-268112P.  
 PR 26-FEB-2001; 2001US-271639P.  
 PR 07-SEP-2001; 2001US-317818P.  
 PR 21-DEC-2001; 2001US-343553P.  
 XX  
 PA (INCYTE) INCYTE GENOMICS INC.  
 XX  
 PI Tang TY, Yue H, Gandhi AR, Yao MG, Warren BA, Ding L, Duggan BW,  
 PI Xu Y, Yang J, Thangavelu K, Lal PG, Honchell CD, Walla NK, Lee S;  
 PI Lee EA, Richardson TW, Baughn MR, Elliott VS;  
 XX WPI; 2002-657522/70.  
 DR N-PSDB; ABS64954.  
 XX  
 PT New human secreted proteins and nucleic acids useful in diagnosing,  
 PT treating and preventing cell proliferative, autoimmune/inflammatory,  
 PT cardiovascular, neurological, and developmental disorders -  
 XX  
 PS Claim 1; Page 140; 158pp; English.

CC The invention relates to twenty four human secreted proteins  
 CC (SECP1-24), proteins 90% identical to them and active fragments of them.  
 CC Also included are nucleic acids encoding the SECP proteins, a recombinant  
 CC polynucleotide comprising a promoter sequence operably linked to the  
 CC nucleic acid, a cell transformed with the recombinant polynucleotide, an  
 CC transgenic organism comprising the recombinant polynucleotide, an  
 CC anti-SECP antibody, and screening for ant/agonists and modulators of  
 CC SECP function or expression. The SECP proteins and nucleic acids are  
 CC useful in the diagnosis, treatment and prevention of cell proliferative  
 CC (e.g. actinic keratosis, arteriosclerosis, bursitis, hepatitis or  
 CC cancer), autoimmune/inflammatory (e.g. AIDS (acquired immunodeficiency  
 CC syndrome), asthma, anaemia, allergies or atopic dermatitis),  
 CC cardiovascular (e.g. congestive heart failure, ischaemic heart disease,  
 CC myocardial infarction, hypertensive heart disease, or vascular tumours),  
 CC neurological (e.g. epilepsy, stroke, cerebral neoplasms, or Alzheimer's  
 CC disease), and developmental (e.g. renal tubular acidosis, Cushing's  
 CC syndrome, Duchenne and Becker muscular dystrophy, or hypothyroidism)  
 CC disorders. Many other diseases and disorders are listed in the  
 CC specification. These may also be used in assessing the effects of  
 CC exogenous compounds on the expression of nucleic acid and amino acid  
 CC sequences of the secreted proteins. The present sequence represents a  
 CC SECP protein of the invention.

SQ Sequence 329 AA;

Query Match 100.0%; Score 894; DB 23; Length 329;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-89;

Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADADAPARGPAPPEPRSAFSAARTSLVSGDAGPGRHOPLAEDTEFVNIGDPDAAAG 60  
 DB 160 ADADAPARGPAPPEPRSAFSAARTSLVSGDAGPGRHOPLAEDTEFVNIGDPDAAAG 219

QY 61 VFRCRLPGAYFFSFTLGLPKRTLSVKLMKRDVQAMITYDDGASRRREMOSQSVMLALR 120  
 DB 220 VFRCRLPGAYFFSFTLGLPKRTLSVKLMKRDVQAMITYDDGASRRREMOSQSVMLALR 279

QY 121 RGDVWMLSHDHGAGYASNHGKTYTFSGFLVYDPLAPAPPGIGASELL 170  
 DB 280 RGDVWMLSHDHGAGYASNHGKTYTFSGFLVYDPLAPAPPGIGASELL 329

RESULT 3

AAAB61424  
 ID AAAB61424 standard; Protein; 329 AA.

AC AAAB61424;

DT 04-APR-2001 (first entry)

XX Monkey MANGO 245 protein.

KM TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
KM autoimmunity; allergy; cardiovascular; brain; degenerative; placental;  
KM pancreatic; skeletal; muscle.  
XX  
XX Catarhini sp..  
XX  
XX WO200100672-A1.  
XX  
XX PD 04-JAN-2001.  
XX  
XX PF 29-JUN-2000; 2000WO-US18184.  
XX  
XX PR 29-JUN-1999; 99US-0342687.  
XX  
XX PA (MILL-) MILLENNIUM PHARM INC.  
XX  
XX PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;  
XX WPI; 2001-050127/06.  
XX  
XX PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
XX PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
XX PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
XX PT disorders (e.g. jaundice) -  
XX  
XX PS Claim 1; Fig 25; 262pp; English.  
XX  
XX CC The present invention relates to cDNAs encoding TANGO 244,  
XX CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
XX CC The nucleic acids, proteins and protein modulators are useful for  
XX CC treating colonic disorders, inflammatory diseases, tumors,  
XX CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
XX CC allergic diseases, cardiovascular diseases, brain disorders,  
XX CC degenerative diseases placental, pancreatic, skeletal and muscle  
XX CC disorders.  
XX  
XX SQ Sequence 329 AA;  
XX  
XX Query Match 97.4%; Score 871; DB 22; Length 329;  
XX Best Local Similarity 97.6%; Pred. No. 1.3e-86;  
XX Matches 166; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
XX  
XX QY 1 ADADAAPARGPAPPEPRSAFSAARTSLVSGDAGPGPRHOPLAFDTEFVNIGDFPAAAG 60  
XX DB 160 ADADAAPARGPAPPEPRSAFSAARTSLVSGDAGPGPRHOPLAFDTEFVNIGDFPAAAG 219  
XX  
XX QY 61 VFRCLPGAYFFSFTLGKLPKRTLSVKLMKNRDEVQAMTYDDGASRRREMOSQSVMLALR 120  
XX DB 220 VFRCLPGAYFFSFTLGKLPKRTLSVKLMKNRDEVQAMTYDDGASRRREMOSQSVMLALR 279  
XX  
XX QY 121 RGDVWLLSHDHGQYGAYSNHGKYITFSGLVYPDLAPAPPGIGASELL 170  
XX DB 280 RGDVWLLSHDHGQYGAYSNHGKYITFSGLVYPDLAPAPPGIGASELL 329  
XX  
XX  
XX RESULT 4  
XX ID AAB61479 standard; protein; 192 AA.  
XX  
XX AC AAB61479;  
XX  
XX DT 04-APR-2001 (first entry)  
XX  
XX DE Human MANGO 245 cytoplasmic domain.  
XX  
XX KM TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
XX KM autoimmunity; allergy; cardiovascular; brain; degenerative; placental;  
XX KM pancreatic; skeletal; muscle.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200100672-A1.  
XX

PD 04-JAN-2001.  
XX  
XX PF 29-JUN-2000; 2000WO-US18184.  
XX  
XX PR 29-JUN-1999; 99US-0342687.  
XX  
XX PA (MILL-) MILLENNIUM PHARM INC.  
XX  
XX PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;  
XX WPI; 2001-050127/06.  
XX  
XX PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
XX PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
XX PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
XX PT disorders (e.g. jaundice) -  
XX  
XX PS Disclosure; Page 247; 262pp; English.  
XX  
XX CC The present invention relates to cDNAs encoding TANGO 244,  
XX CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
XX CC The nucleic acids, proteins and protein modulators are useful for  
XX CC treating colonic disorders, inflammatory diseases, tumors,  
XX CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
XX CC allergic diseases, cardiovascular diseases, brain disorders,  
XX CC degenerative diseases placental, pancreatic, skeletal and muscle  
XX CC disorders.  
XX  
XX SQ Sequence 192 AA;  
XX  
XX Query Match 83.1%; Score 743; DB 22; Length 192;  
XX Best Local Similarity 100.0%; Pred. No. 6.2e-73;  
XX Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 1 ADADAAPARGPAPPEPRSAFSAARTSLVSGDAGPGPRHOPLAFDTEFVNIGDFPAAAG 60  
XX DB 4 ADADAAPARGPAPPEPRSAFSAARTSLVSGDAGPGPRHOPLAFDTEFVNIGDFPAAAG 63  
XX  
XX QY 61 VFRCLPGAYFFSFTLGKLPKRTLSVKLMKNRDEVQAMTYDDGASRRREMOSQSVMLALR 120  
XX DB 64 VFRCLPGAYFFSFTLGKLPKRTLSVKLMKNRDEVQAMTYDDGASRRREMOSQSVMLALR 123  
XX  
XX QY 121 RGDVWLLSHDHGQYGAYSNH 141  
XX DB 124 RGDVWLLSHDHGQYGAYSNH 144  
XX  
XX  
XX RESULT 5  
XX ID AAB61466 standard; protein; 334 AA.  
XX  
XX AC AAB61466;  
XX  
XX DT 04-APR-2001 (first entry)  
XX  
XX DE Human MANGO 245 mature protein.  
XX  
XX KM TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
XX KM autoimmunity; allergy; cardiovascular; brain; degenerative; placental;  
XX KM pancreatic; skeletal; muscle.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200100672-A1.  
XX  
XX PD 04-JAN-2001.  
XX  
XX PF 29-JUN-2000; 2000WO-US18184.  
XX  
XX PR 29-JUN-1999; 99US-0342687.  
XX  
XX PA (MILL-) MILLENNIUM PHARM INC.  
XX

PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;  
XX WPI; 2001-050127/06.  
XX  
XX Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
PT disorders (e.g. jaundice) -  
XX  
XX Disclosure; Page 242; 262pp; English.  
XX  
XX The present invention relates to cDNAs encoding TANGO 244,  
CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
CC The nucleic acids, proteins and protein modulators are useful for  
CC treating colonic disorders, inflammatory diseases, tumors,  
CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
CC allergic diseases, cardiovascular diseases, brain disorders,  
CC degenerative diseases placental, pancreatic, skeletal and muscle  
CC disorders.  
CC  
SQ Sequence 334 AA;  
Query Match 83.1%; Score 743; DB 22; Length 334;  
Best Local Similarity 100.0%; Pred. No. 1.2e-72;  
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ADADAPARGPPAPPEPRSAFSAARTSLVGSDDAGPPRHQPLAFDTEFVNIGDPPAAAG 60  
DB 146 ADADAPARGPPAPPEPRSAFSAARTSLVGSDDAGPPRHQPLAFDTEFVNIGDPPAAAG 205  
QY 61 VFCRLPGAYFFSFTLGLPKRTLSVKLMKRDVQAMITYDDGASRRRMOSSQVMLALR 120  
DB 206 VFCRLPGAYFFSFTLGLPKRTLSVKLMKRDVQAMITYDDGASRRRMOSSQVMLALR 265  
QY 121 RGDVWMLSHDHGAYGAYSNH 141  
DB 266 RGDVWMLSHDHGAYGAYSNH 266  
RESULT 6  
AAB61423  
ID AAB61423 standard; protein; 348 AA.  
XX  
XX AAB61423;  
AC  
XX 04-APR-2001 (first entry)  
DT  
XX  
XX Human MANGO 245 protein.  
DE  
XX TANGO, MANGO, colon; inflammation; tumor; renal; liver; lung;  
KM autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
KM pancreatic; skeletal; muscle.  
XX  
XX Homo sapiens.  
OS  
XX WO200100672-A1.  
PN  
XX 04-JAN-2001.  
PD  
XX 29-JUN-2000; 2000WO-US18184.  
PF  
XX 29-JUN-1999; 99US-0342687.  
PR  
XX (MILL-) MILLENNIUM PHARM INC.  
PA  
XX Holtzman DA, Barnes TM, Fraser CC, Sharp JD;  
PI WPI; 2001-050127/06.  
XX  
XX Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
PT disorders (e.g. jaundice) -

XX  
XX Claim 1; Fig 23; 262pp; English.  
PS  
XX  
XX The present invention relates to cDNAs encoding TANGO 244,  
CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
CC The nucleic acids, proteins and protein modulators are useful for  
CC treating colonic disorders, inflammatory diseases, tumors,  
CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
CC allergic diseases, cardiovascular diseases, brain disorders,  
CC degenerative diseases placental, pancreatic, skeletal and muscle  
CC disorders.  
CC  
SQ Sequence 348 AA;  
Query Match 83.1%; Score 743; DB 22; Length 348;  
Best Local Similarity 100.0%; Pred. No. 1.3e-72;  
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ADADAPARGPPAPPEPRSAFSAARTSLVGSDDAGPPRHQPLAFDTEFVNIGDPPAAAG 60  
DB 160 ADADAPARGPPAPPEPRSAFSAARTSLVGSDDAGPPRHQPLAFDTEFVNIGDPPAAAG 219  
QY 61 VFCRLPGAYFFSFTLGLPKRTLSVKLMKRDVQAMITYDDGASRRRMOSSQVMLALR 120  
DB 220 VFCRLPGAYFFSFTLGLPKRTLSVKLMKRDVQAMITYDDGASRRRMOSSQVMLALR 279  
QY 121 RGDVWMLSHDHGAYGAYSNH 141  
DB 280 RGDVWMLSHDHGAYGAYSNH 300  
RESULT 7  
ABG70385  
ID ABG70385 standard; protein; 299 AA.  
XX  
XX ABG70385;  
AC  
XX 05-NOV-2002 (first entry)  
DT  
XX  
XX Adipocyte complement-related C1q Tumour Necrosis Factor-like protein #3.  
DE  
XX Human; NOX; NOX-associated disorder; cardiomyopathy; atherosclerosis;  
KM cell signal processing; metabolic pathway modulation; metabolic disorder;  
KM obesity; diabetes; infectious disease; neurodegenerative disorder; acne;  
KM Alzheimer's disease; Parkinson's disease; immune disorder; cancer;  
KM haematopoietic disorder; cirrhosis; pancreatitis; learning defect;  
KM memory defect; infertility; congenital heart defect; hair growth;  
KM pigmentation disorder; endocrine disorder; respiratory disease; health;  
KM gastro-intestinal disease; reproductive; neurological disease;  
KM bone marrow transplantation; endocrine disease; allergy; inflammation;  
KM nephrological disorder; urinary system disorder; age-related disorder;  
KM neuropsychiatric disorder; EGF-related protein; SCUB1; TEN-M4;  
KM adipocyte complement-related C1q tumour necrosis factor; out at first;  
KM beta adrenergic receptor kinase; EphA6/ehk-2; glucose transporter;  
KM type 1a membrane sushi-containing domain; butyrophilin;  
XX type 1a membrane-sushi domain containing.  
OS  
XX Homo sapiens.  
XX  
XX WO200257453-A2.  
PN  
XX 25-JUL-2002.  
PD  
XX 19-DEC-2001; 2001WO-US50331.  
PF  
XX 19-DEC-2000; 2000US-265704P.  
PR 19-DEC-2000; 2000US-257314P.  
PR 02-MAY-2001; 2001US-288153P.  
PR 29-MAY-2001; 2001US-294075P.  
PR 24-JUL-2001; 2001US-307506P.  
PR 10-AUG-2001; 2001US-311590P.  
PR 10-AUG-2001; 2001US-311613P.  
PR 29-AUG-2001; 2001US-315617P.

PR 14-SEP-2001; 2001US-322358P.  
XX (CURA-) CUDAGEN CORP.  
XX  
XX  
PI Gangolli EA, Patturajan M, Vernet CAM, Malyanekar UM, Kekuda R;  
PI Stone DJ, Anderson D, Shinkets RA, Burgess CE, Zernusen BD, Liu X;  
PI Spytek KA, Casman SJ, Boldog FL, Smithson G, Li L, Ji W;  
XX MPI: 2002-590744/63.  
DR N-PSDB; ABS52097.  
XX  
XX Novel isolated NOVX polypeptide useful for treating cardiomyopathy,  
PT atherosclerosis, metabolic disorders, diabetes, obesity, infectious  
PT disease, anorexia, neurodegenerative disorders, Alzheimer's disease or  
PT cancer.  
XX  
XX Claim 1; Page 31; 318pp; English.  
XX  
XX The present invention relates to new NOVX polypeptides. The invention is  
XX useful for treating or preventing a NOVX-associated disorder such as  
XX cardiomyopathy or atherosclerosis, where the disorder is related to cell  
XX signal processing and metabolic pathway modulation in a subject,  
XX preferably human. The invention is also useful for treating metabolic  
XX disorders (e.g. obesity), diabetes, infectious disease, neurodegenerative  
XX disorders (e.g. Alzheimer's disease, Parkinson's disease), immune  
XX disorders, haematopoietic disorders and various cancers. The molecules of  
XX the invention are also useful for treating or preventing cirrhosis,  
XX pancreatitis, learning and memory defects, infertility, congenital heart  
XX defects, acne, hair growth, pigmentation disorders, endocrine disorders,  
XX respiratory disease, gastro-intestinal diseases, reproductive health,  
XX neurological diseases, bone marrow transplantation, endocrine diseases,  
XX allergy and inflammation, nephrological disorders, urinary system  
XX disorders, neuropsychiatric disorders and age-related disorders.  
XX The present amino acid sequence represents a NOVX protein of the  
XX invention.  
XX  
XX Sequence 299 AA:  
SQ  
Query Match 81.7%; Score 730; DB 23; Length 299;  
Best Local Similarity 98.6%; Pred. No. 2.9e-71;  
Matches 138; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 31 SDAGGPRHOPAPFTFFVNIIGDDPDAAGVRCRLPGAYFFSFTLGKLPKRTLSVKLMK 90  
DB 160 ADAGGPRHOPAPFTFFVNIIGDDPDAAGVRCRLPGAYFFSFTLGKLPKRTLSVKLMK 219  
QY 91 NRDEVOAMITDYGASRRREMOSQSVMLALRQDAVWMLSHDHDGAYSNHGKTYTFSGF 150  
DB 220 NRDEVOAMITDYGASRRREMOSQSVMLALRQDAVWMLSHDHDGAYSNHGKTYTFSGF 279  
QY 151 LVYPDLAPAPPGGICASELL 170  
DB 280 LVYPDLAPAPPGGICASELL 299  
RESULT 8  
ABG70383 ID ABG70383 standard; Protein; 221 AA.  
XX  
XX ABG70383;  
AC  
XX  
XX 05-NOV-2002 (first entry)  
XX  
XX Adipocyte complement-related C1q Tumour Necrosis Factor-like protein #1.  
XX  
XX Human; NOVX; NOVX-associated disorder; cardiomyopathy; atherosclerosis;  
XX cell signal processing; metabolic pathway modulation; metabolic disorder;  
XX obesity; diabetes; infectious disease; neurodegenerative disorder; acne;  
XX Alzheimer's disease; Parkinson's disease; immune disorder; cancer;  
XX haematopoietic disorder; cirrhosis; pancreatitis; learning defect;  
XX memory defect; infertility; congenital heart defect; hair growth;  
XX pigmentation disorder; endocrine disorder; respiratory disease; health;  
XX gastro-intestinal disease; reproductive; neurological disease;

KW bone marrow transplantation; endocrine disease; allergy; inflammation;  
KW neurological disorder; urinary system disorder; age-related disorder;  
KW neuropsychiatric disorder; EGF-related protein; SCUB2; TEN-M4;  
KW adipocyte complement-related C1q tumour necrosis factor; out at first;  
KW beta adrenergic receptor kinase; EphA6/btk-2; glucose transporter;  
KW type Ia membrane sueshi-containing domain; butyrophilin.  
XX  
XX Homo sapiens.  
XX  
XX WO200257453-A2.  
XX  
XX 25-JUL-2002.  
XX  
XX 19-DEC-2001; 2001WO-US50331.  
XX  
XX 19-DEC-2000; 2000US-265704P.  
XX 20-DEC-2000; 2000US-257314P.  
XX 02-MAY-2001; 2001US-286153P.  
XX 29-MAY-2001; 2001US-294075P.  
XX 24-JUL-2001; 2001US-307506P.  
XX 10-AUG-2001; 2001US-311590P.  
XX 10-AUG-2001; 2001US-311613P.  
XX 29-AUG-2001; 2001US-315617P.  
XX 14-SEP-2001; 2001US-322358P.  
XX  
XX (CURA-) CUDAGEN CORP.  
XX  
XX Gangolli EA, Patturajan M, Vernet CAM, Malyanekar UM, Kekuda R;  
XX Stone DJ, Anderson D, Shinkets RA, Burgess CE, Zernusen BD, Liu X;  
XX Spytek KA, Casman SJ, Boldog FL, Smithson G, Li L, Ji W;  
XX MPI: 2002-590744/63.  
XX N-PSDB; ABS52095.  
XX  
XX Novel isolated NOVX polypeptide useful for treating cardiomyopathy,  
XX atherosclerosis, metabolic disorders, diabetes, obesity, infectious  
XX disease, anorexia, neurodegenerative disorders, Alzheimer's disease or  
XX cancer.  
XX  
XX Claim 1; Page 28; 318pp; English.  
XX  
XX The present invention relates to new NOVX polypeptides. The invention is  
XX useful for treating or preventing a NOVX-associated disorder such as  
XX cardiomyopathy or atherosclerosis, where the disorder is related to cell  
XX signal processing and metabolic pathway modulation in a subject,  
XX preferably human. The invention is also useful for treating metabolic  
XX disorders (e.g. obesity), diabetes, infectious disease, neurodegenerative  
XX disorders (e.g. Alzheimer's disease, Parkinson's disease), immune  
XX disorders, haematopoietic disorders and various cancers. The molecules of  
XX the invention are also useful for treating or preventing cirrhosis,  
XX pancreatitis, learning and memory defects, infertility, congenital heart  
XX defects, acne, hair growth, pigmentation disorders, endocrine disorders,  
XX respiratory disease, gastro-intestinal diseases, reproductive health,  
XX neurological diseases, bone marrow transplantation, endocrine diseases,  
XX allergy and inflammation, nephrological disorders, urinary system  
XX disorders, neuropsychiatric disorders and age-related disorders.  
XX The present amino acid sequence represents a NOVX protein of the  
XX invention.  
XX  
XX Sequence 221 AA:  
SQ  
Query Match 77.9%; Score 696; DB 23; Length 221;  
Best Local Similarity 76.1%; Pred. No. 1e-67;  
Matches 140; Conservative 3; Mismatches 13; Indels 28; Gaps 2;  
QY 9 GPPAP-----EPRSAFSAARTSLVGSAGPGRHOPAPFT 46  
DB 44 GPPFPGLPKRGRAGLSGKNGFPDGSASRARSASARTPLEGTS-----EWAIVFDK 97  
QY 47 EFVNIIGDDPDAAGVRCRLPGAYFFSFTLGKLPKRTLSVKLMNRDEVOAMITDYGASR 106  
DB 98 VVNIIGDDPDAAGVRCRLPGAYFFSFTLGKLPKRTLSVKLMNRDEVOAMITDYGASR 157

QY 107 REMOSQSVMLALRRGDVAVMLSHDHDGYGAYSNHGKXITTFSGFLVYPDLPAPAPGLGA 166  
 DB 158 REMOSQSVMLALRRGDVAVMLSHDHDGYGAYSNHGKXITTFSGFLVYPDLPAPAPGLGA 217  
 QY 167 SEEL 170  
 DB 218 SEEL 221

RESULT 9

AAB61473  
 ID AAB61473 standard; protein; 134 AA.  
 XX  
 AC AAB61473;  
 DT 04-APR-2001 (first entry)  
 XX  
 DE Monkey MANGO 245 C1q domain #2.  
 XX  
 KW TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
 KW autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
 KW pancreatic; skeletal; muscle.  
 XX  
 OS Catarrhini sp.  
 XX  
 PN MO200100672-A1.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 29-JUN-2000; 2000MO-US18184.  
 XX  
 PR 29-JUN-1999; 99US-0342687.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Holzman DA, Barnes TM, Fraser CC, Sharp JD;  
 XX  
 DR WPI; 2001-050127/06.  
 XX  
 PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
 PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
 PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
 PT disorders (e.g. jaundice) -  
 XX  
 PS Disclosure; Fig 28; 262pp; English.  
 XX  
 CC The present invention relates to cDNAs encoding TANGO 244,  
 CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
 CC The nucleic acids, proteins and protein modulators are useful for  
 CC treating colonic disorders, inflammatory diseases, tumors,  
 CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
 CC allergic diseases, cardiovascular diseases, brain disorders,  
 CC degenerative diseases placental, pancreatic, skeletal and muscle  
 CC disorders.  
 CC  
 SQ Sequence 134 AA;

Query Match 76.4%; Score 683; DB 22; Length 134;  
 Best Local Similarity 97.8%; Pred. No. 1.4e-66;  
 Matches 131; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 19 AFSARTSTSVGSDGPGPRHQPLAFDTEFVNIGGDFPAAAGVFRCLPGAYFFSFTLQK 78  
 DB 1 AFSARTSTSVGSDGPGPRHQPLAFDTEFVNIGGDFPAAAGVFRCLPGAYFFSFTLQK 60  
 QY 79 LPRKTLVYLMKNRDEVOQAMIVDDASRRREMOSQSVMLALRRGDVAVMLSHDHDGYGAY 138  
 DB 61 LPRKTLVYLMKNRDEVOQAMIVDDASRRREMOSQSVMLALRRGDVAVMLSHDHDGYGAY 120  
 QY 139 SNHGKXITTFSGFLV 152  
 DB 121 SNHGKXITTFSGFLV 134

RESULT 10

ABG70384  
 ID ABG70384 standard; Protein; 284 AA.  
 XX  
 AC ABG70384;  
 DT 05-NOV-2002 (first entry)  
 XX  
 DE Adipocyte complement-related C1q Tumour Necrosis Factor-like protein #2.

XX  
 KW Human; NOX; NOX-associated disorder; cardiomyopathy; atherosclerosis;  
 KW cell signal processing; metabolic pathway modulation; metabolic disorder;  
 KW obesity; diabetes; infectious disease; neurodegenerative disorder; acne;  
 KW Alzheimer's disease; Parkinson's disease; immune disorder; cancer;  
 KW haematopoietic disorder; cirrhosis; pancreatitis; learning defect;  
 KW memory defect; infertility; congenital heart defect; hair growth;  
 KW pigmentation disorder; endocrine disorder; respiratory disease; health;  
 KW gastro-intestinal disease; reproductive; neurological disease;  
 KW bone marrow transplantation; endocrine disease; allergy; inflammation;  
 KW nephrological disorder; urinary system disorder; age-related disorder;  
 KW neuropsychiatric disorder; EGF-related protein; SCUBE; TEN-4;  
 KW adipocyte complement-related C1q tumour necrosis factor; cut at first;  
 KW beta adrenergic receptor kinase; EphA6/ehk-2; glucose transporter;  
 KW type 1a membrane-sushi-containing domain; butyrophillin;  
 KW type 1a membrane-sushi-containing domain.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200257453-A2.  
 XX  
 PD 25-JUL-2002.  
 XX  
 PF 19-DEC-2001; 2001MO-US50331.  
 XX  
 PR 19-DEC-2000; 2000US-265704P.  
 PR 20-DEC-2000; 2000US-257314P.  
 PR 02-MAY-2001; 2001US-288153P.  
 PR 29-MAY-2001; 2001US-294075P.  
 PR 24-JUL-2001; 2001US-307506P.  
 PR 10-AUG-2001; 2001US-311593P.  
 PR 10-AUG-2001; 2001US-311613P.  
 PR 29-AUG-2001; 2001US-315617P.  
 PR 14-SEP-2001; 2001US-322358P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Ganggoli EA, Patuvarajan M, Vernet CM, Maliyankar UM, Kekuda R,  
 PI Stone DJ, Anderson D, Shimkets RA, Burgess CE, Zernusen BD, Liu X,  
 PI Szytek KA, Casman SJ, Boldog FL, Smithson G, Li L, Ji W;  
 XX  
 DR WPI; 2002-590744/63.  
 DR N-PSDB; ABS52096.  
 XX  
 PT Novel isolated NOX polypeptide useful for treating cardiomyopathy,  
 PT atherosclerosis, metabolic disorders, diabetes, obesity, infectious  
 PT disease, anorexia, neurodegenerative disorders, Alzheimer's disease or  
 PT cancer -  
 XX  
 PS Claim 1; Page 30; 318pp; English.  
 XX  
 CC The present invention relates to new NOX polypeptides. The invention is  
 CC useful for treating or preventing a NOX-associated disorder such as  
 CC cardiomyopathy or atherosclerosis, where the disorder is related to cell  
 CC signal processing and metabolic pathway modulation in a subject,  
 CC preferably human. The invention is also useful for treating metabolic  
 CC disorders (e.g. obesity, diabetes, infectious disease, neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease), immune  
 CC disorders, haematopoietic disorders and various cancers. The molecules of  
 CC the invention are also useful for treating or preventing cirrhosis,  
 CC pancreatitis, learning and memory defects, infertility, congenital heart  
 CC defects, acne, hair growth, pigmentation disorders, endocrine disorders,

CC respiratory disease, gastro-intestinal diseases, reproductive, health.  
 CC neurological diseases, bone marrow transplantation, endocrine diseases,  
 CC allergy and inflammation, nephrological disorders, urinary system  
 CC disorders, neuropsychiatric disorders and age-related disorders.  
 CC The present amino acid sequence represents a NOVX protein of the  
 CC invention.

XX  
 SQ Sequence 284 AA;

Query Match 74.6%; Score 666.5; DB 23; Length 284;  
 Best Local Similarity 55.4%; Pred. No. 2.3e-64;  
 Matches 149; Conservative 1; Mismatches 14; Indels 105; Gaps 5;

QY 7 ARGP---PAPPEPSAFAAATRSI-----VGSD----- 32  
 DB 16 ALGPTPGSGSELSSAFSAARTPLTEGTSENAVTFDKYVYVIGDFVATGQFCRVYGA 75  
 QY 33 -----AGPGRH-----OPLAF----- 44  
 DB 76 YFFSFTACKAPHKLSVWLVNRDEVMALRDEQRPGARRAASQASMLQLDYGDTWLR 135  
 QY 45 -----DTEFVNIIGDFDAAGVFRCLPGAYFFSFTLGKLP 81  
 DB 136 LHGAPOVALGAPATFSGYLVYADAEPVNIIGDFDAAGVFRCLPGAYFFSFTLGKLP 195  
 QY 82 KTLSEVKLMKNRDEVOAMTYDGGASRRRMOQSVMALRRGDVWLSHHDGAYSNH 141  
 DB 196 KTLSEVKLMKNRDEVOAMTYDGGASRRRMOQSVMALRRGDVWLSHHDGAYSNH 255  
 QY 142 GKTYTFSGFLVYPDLAPAPPGGLGSELL 170  
 DB 256 GKTYTFSGFLVYPDLAPAPPGGLGSELL 284

RESULT 11  
 ABB1578  
 ID ABB1578 standard; peptide: 127 AA.

XX  
 DT 11-JAN-2002 (first entry)

DE Human secreted protein homologue, SEQ ID NO:1948.

XX Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
 KW chronic inflammatory condition; proliferative retinopathy;  
 KW atherosclerosis; coronary heart disease; arterial ischaemia;  
 KW bone disorder; osteoporosis; vascular growth disorder;  
 KW tissue regeneration; wound healing; infection; immune disorder;  
 KW cell culture; drug screening; gene therapy; antiinflammatory;  
 KW antiasthmatic; antiarthritic; haemostatic; antileukosclerotic;  
 KW cytoskeletal; osteoplastic; vasotropic; cardiant; virucide; antibacterial;  
 KW antifungal; vulnery; antiulcer.

XX Homo sapiens.

OS  
 PN W0200157188-A2.

XX  
 PD 09-AUG-2001.

XX  
 PF 05-FEB-2001; 2001WO-US03800.

XX  
 PR 03-FEB-2000; 2000US-0496914.

XX  
 PR 27-APR-2000; 2000US-0560875.

XX  
 PA (HYSE-) HYSEQ INC.

XX  
 PI Tang YT, Liu C, Drmanac RT;

DR WPI: 2001-457740/49.

DR N-PSDB; ABA08822.

PT Human proteins and DNA encoding sequences useful for preventing,  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer -  
 XX  
 PS Claim 20; Page 214-215; 1963pp; English.

XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
 CC sequences ABA08822-ABA09574 represent nucleic acids encoding them. The  
 CC invention also relates to vectors and recombinant host cells comprising a  
 CC nucleotide of the invention, methods of producing the novel polypeptides,  
 CC antibodies against the polypeptides, methods of detecting the nucleotides  
 CC or polypeptides in a sample, and methods of identifying compounds which  
 CC bind to polypeptides of the invention. Although novel, many of the  
 CC polypeptides of the invention have homology to known proteins, thereby  
 CC giving an insight into their probable biological activities, and hence  
 CC potential therapeutic applications. The polypeptides of the invention may  
 CC have various activities, including cytokine, cell proliferation or cell  
 CC differentiation activities; stem cell growth factor activity;  
 CC haematopoiesis regulatory activity; tissue growth factor activity;  
 CC immunomodulatory activity; activin- or inhibin-related activities;  
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 CC thrombolytic activities; receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC Depending on their biological activities, polypeptides and nucleotides of  
 CC the invention are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC vascular growth. Polypeptides involved with tissue regeneration and  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders.  
 CC Polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells  
 CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a novel human  
 CC polypeptide of the invention.

XX  
 SQ Sequence 127 AA;

Query Match 74.3%; Score 664; DB 22; Length 127;  
 Best Local Similarity 99.2%; Pred. No. 1.6e-64;  
 Matches 126; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 44 FDFEFVNIIGDFDAAGVFRCLPGAYFFSFTLGKLPKRLSYKLMKNRDEVOAMTYDDG 103

DB 1 FDFEFVNIIGDFDAAGVFRCLPGAYFFSFTLGKLPKRLSYKLMKNRDEVOAMTYDDG 60

QY 104 ASRRRMOQSVMALRRGDVWLSHHDGAYSNHGKTYTFSGFLVYPDLAPAPPG 163

DB 61 SSRRRMOQSVMALRRGDVWLSHHDGAYSNHGKTYTFSGFLVYPDLAPAPPG 120

QY 164 LGASELL 170

DB 121 LGASELL 127

RESULT 12

AAB61470

ID AAB61470 standard; protein; 117 AA.

XX  
 AC AAB61470;

XX  
 DT 04-APR-2001 (first entry)

XX DE Human MANGO 245 C1q domain #2.  
XX XX  
XX TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
KW autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
KW pancreatic; skeletal; muscle.  
XX OS Homo sapiens.  
XX XX  
XX PN WO200100672-A1.  
XX XX  
XX PD 04-JAN-2001.  
XX XX  
XX PF 29-JUN-2000; 2000WO-US18184.  
XX XX  
XX PR 29-JUN-1999; 99US-0342687.  
XX XX  
XX PA (MILL-) MILLENNIUM PHARM INC.  
XX XX  
XX PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD.  
XX DR WPI; 2001-050127/06.  
XX XX  
XX PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
XX disorders (e.g. jaundice) -  
XX XX  
XX PS Disclosure; Fig 26; 262pp; English.  
XX XX  
XX CC The present invention relates to cDNAs encoding TANGO 244,  
XX CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
XX CC The nucleic acids, proteins and protein modulators are useful for  
XX CC treating colonic disorders, inflammatory diseases, tumors,  
XX CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
XX CC allergic diseases, cardiovascular diseases, brain disorders,  
XX CC degenerative diseases placental, pancreatic, skeletal and muscle  
XX CC disorders.  
XX XX  
XX SQ Sequence 117 AA;  
XX XX  
XX Query Match 68.1%; Score 609; DB 22; Length 117;  
XX Best Local Similarity 100.0%; Pred. No. 1.5e-58;  
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX XX  
XX QY 19 AFSARTSLVGSDDAGPGRHOPLAFTPEFVNIGDPPAAGVPRCRLPGAYFSGFTLGG 78  
XX DB 1 AFSARTSLVGSDDAGPGRHOPLAFTPEFVNIGDPPAAGVPRCRLPGAYFSGFTLGG 60  
XX QY 79 LPRKTLVYLMKNRDEVOAMITYDGASRRRMOQSVMALRRGDVWLLSHDHG 135  
XX DB 61 LPRKTLVYLMKNRDEVOAMITYDGASRRRMOQSVMALRRGDVWLLSHDHG 117  
XX XX  
XX RESULT 13  
XX AAB61468  
XX ID AAB61468 standard; protein; 130 AA.  
XX XX  
XX AC AAB61468;  
XX XX  
XX DT 04-APR-2001 (first entry)  
XX XX  
XX DE Mature monkey MANGO 245 protein.  
XX XX  
XX KW TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
KW autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
KW pancreatic; skeletal; muscle.  
XX OS Catarhini sp..  
XX XX  
XX PN WO200100672-A1.  
XX XX  
XX PD 04-JAN-2001.  
XX XX

XX XX  
XX PF 29-JUN-2000; 2000WO-US18184.  
XX XX  
XX PR 29-JUN-1999; 99US-0342687.  
XX XX  
XX PA (MILL-) MILLENNIUM PHARM INC.  
XX XX  
XX PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;  
XX XX  
XX DR WPI; 2001-050127/06.  
XX XX  
XX PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
XX disorders (e.g. jaundice) -  
XX XX  
XX PS Disclosure; Page 243; 262pp; English.  
XX XX  
XX CC The present invention relates to cDNAs encoding TANGO 244,  
XX CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
XX CC The nucleic acids, proteins and protein modulators are useful for  
XX CC treating colonic disorders, inflammatory diseases, tumors,  
XX CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
XX CC allergic diseases, cardiovascular diseases, brain disorders,  
XX CC degenerative diseases placental, pancreatic, skeletal and muscle  
XX CC disorders.  
XX XX  
XX SQ Sequence 130 AA;  
XX XX  
XX Query Match 53.9%; Score 482; DB 22; Length 130;  
XX Best Local Similarity 61.2%; Pred. No. 1.3e-44;  
XX Matches 101; Conservative 6; Mismatches 22; Indels 36; Gaps 3;  
XX XX  
XX QY 6 PARPPAPPRPSAFSARTSLVGSDDAGPGRHOPLAFTPEFVNIGDPPAAGVPRCR 65  
XX DB 2 PARPPAPPRPSAFSARTSLVGSDDAGPGRHOPLAFTPEFVNIGDPPAAGVPRCR 54  
XX QY 66 LPGAFFSFTLGLPRKTLVYLMKNRDEVOAMITYDGASRRRMOQSVMALRRGDV 125  
XX DB 55 VPGAFFSFTLGLPRKTLVYLMKNRDEVOAMITYDGASRRRMOQSVMALRRGDV 85  
XX QY 126 WLSHHDHGGYAVSNHGKTYITFSGFLVYPPDLAPAPGLGASLL 170  
XX DB 86 WLSHHDHGGYAVSNHGKTYITFSGFLVYPPDLAPAPGLGASLL 130  
XX XX  
XX RESULT 14  
XX ABG70386  
XX ID ABG70386 standard; Protein; 133 AA.  
XX XX  
XX AC ABG70386;  
XX XX  
XX DT 05-NOV-2002 (first entry)  
XX XX  
XX DE Adipocyte complement-related C1q Tumour Necrosis Factor-like protein #4.  
XX XX  
XX KW Human; NOX; NOX-associated disorder; cardiomyopathy; atherosclerosis;  
KW cell signal processing; metabolic pathway modulation; metabolic disorder;  
KW obesity; diabetes; infectious disease; neurodegenerative disorder; acne;  
KW Alzheimer's disease; Parkinson's disease; immune disorder; cancer;  
KW haematopoietic disorder; cirrhosis; pancreatitis; learning defect;  
KW memory defect; infertility; congenital heart defect; hair growth;  
KW pigmentation disorder; endocrine disorder; respiratory disease; health;  
KW gastro-intestinal disease; reproductive; neurological disease;  
KW bone marrow transplantation; endocrine disease; allergy; inflammation;  
KW nephrological disorder; urinary system disorder; age-related disorder;  
KW neuropsychiatric disorder; EGF-related protein; SCUB1; TEN-M4;  
KW adipocyte complement-related C1q tumour necrosis factor; out at first;  
KW beta adrenergic receptor kinase; EphA6/ehk-2; glucose transporter;  
KW type 1a membrane sushi-containing domain; butyrophilin;  
KW type 1a membrane-sushi domain containing.  
XX OS Homo sapiens.  
XX XX



XX XX MO200257453-A2.  
 PN XX  
 XX XX 25-JUL-2002.  
 PD XX  
 XX XX 19-DEC-2001; 2001WO-US50331.  
 PF XX  
 XX 19-DEC-2000; 2000US-265704P.  
 PR 20-DEC-2000; 2000US-257314P.  
 PR 02-MAY-2001; 2001US-288153P.  
 PR 29-MAY-2001; 2001US-294075P.  
 PR 24-JUL-2001; 2001US-307506P.  
 PR 10-AUG-2001; 2001US-311590P.  
 PR 10-AUG-2001; 2001US-311613P.  
 PR 29-AUG-2001; 2001US-315617P.  
 PR 14-SEP-2001; 2001US-322358P.  
 XX XX  
 PA (CURA-) CURAGEN CORP.  
 PI Gangolli EA, Paturajan M, Vernet CAM, Malpankar UM, Kekuda R,  
 PI Stone DJ, Anderson D, Shimkets RA, Burgess CE, Zehrsen BD, Liu X,  
 PI Spytek KA, Casman SJ, Boldog FL, Smithson G, Li L, Ji W;  
 XX XX  
 DR WPI: 2002-590744/63.  
 DR N-PSDB; ABS52098.  
 XX XX  
 PT Novel isolated NOVX polypeptide useful for treating cardiomyopathy,  
 PT atherosclerosis, metabolic disorders, diabetes, obesity, infectious  
 PT disease, anorexia, neurodegenerative disorders, Alzheimer's disease or  
 PT cancer -  
 XX XX  
 PS Claim 1; Page 32; 318pp; English.  
 XX XX  
 CC The present invention relates to new NOVX polypeptides. The invention is  
 CC useful for treating or preventing a NOVX-associated disorder such as  
 CC cardiomyopathy or atherosclerosis, where the disorder is related to cell  
 CC signal processing and metabolic pathway modulation in a subject,  
 CC preferably human. The invention is also useful for treating metabolic  
 CC disorders (e.g. obesity), diabetes, infectious disease, neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease), immune  
 CC disorders, haematopoietic disorders and various cancers. The molecules of  
 CC the invention are also useful for treating or preventing cirrhosis,  
 CC pancreatitis, learning and memory defects, infertility, congenital heart  
 CC defects, acne, hair growth, pigmentation disorders, endocrine disorders,  
 CC respiratory disease, gastro-intestinal diseases, reproductive, health,  
 CC neurological diseases, bone marrow transplantation, endocrine diseases,  
 CC allergy and inflammation, nephrological disorders, urinary system  
 CC disorders, neuropsychiatric disorders and age-related disorders.  
 CC The present amino acid sequence represents a NOVX protein of the  
 CC invention.  
 XX XX  
 SQ Sequence 133 AA;  
 XX XX  
 Query Match 48.3%; Score 432; DB 22; Length 133;  
 Best Local Similarity 56.3%; Pred. No. 3.8e-35;  
 Matches 94; Conservative 3; Mismatches 18; Indels 52; Gaps 3;  
 QY 7 ARGP---PAPPEPRSAFSAARTSLVGSADGPGPRHOPLAFDTFVNIGDPDAAAGVFR 63  
 Db 16 ALGPPTPGSSSELSASARTTPLEGTS-----EAAVTFDKYYVNIIGDFDATTQCFR 69  
 QY 64 CRLEGAYFFSFTLGKLPKRTLSVTKMKNRDEVQAMIIDGASRRRMOQSQSVMLALRRGD 123  
 Db 70 C-----REMOQSQSVMLALRRGD 86  
 QY 124 AVMLLSHDHOGYGAYSNHGKITYTSGFLVYVYDLAPAAPPGIGASELL 170  
 Db 87 AVMLLSHDHOGYGAYSNHGKITYTSGFLVYVYDLAPAAPPGIGASELL 133  
 RESULT 15  
 AAB61488  
 ID AAB61488 standard; protein; 199 AA.

XX XX AAB61488;  
 AC XX  
 XX XX 04-APR-2001 (first entry)  
 DT XX  
 XX XX Murine MANGO 245 protein.  
 DE XX  
 XX MANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;  
 KW autoimmunity; allergy; cardiovascular; brain; degenerative; placental;  
 KM pancreatic; skeletal; muscle.  
 XX  
 OS Mus musculus.  
 XX  
 PN WO20010672-A1.  
 PN  
 PD 04-JAN-2001.  
 PD  
 XX  
 PF 29-JUN-2000; 2000WO-US18184.  
 PF  
 XX 29-JUN-1999; 99US-0342687.  
 PR  
 XX (MIL-) MILLENNIUM PHARM INC.  
 PA  
 XX  
 PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;  
 PI  
 XX  
 DR WPI: 2001-050127/06.  
 DR  
 XX  
 PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
 PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
 PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
 PT disorders (e.g. jaundice) -  
 XX XX  
 PS Claim 1; Fig 29; 262pp; English.  
 XX XX  
 CC The present invention relates to cDNAs encoding TANGO 244,  
 CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
 CC The nucleic acids, proteins and protein modulators are useful for  
 CC treating colonic disorders, inflammatory diseases, tumors,  
 CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
 CC allergic diseases, cardiovascular diseases, brain disorders,  
 CC degenerative diseases placental, pancreatic, skeletal and muscle  
 CC disorders.  
 XX XX  
 SQ Sequence 199 AA;  
 XX XX  
 Query Match 40.0%; Score 357.5; DB 22; Length 199;  
 Best Local Similarity 52.1%; Pred. No. 8.7e-31;  
 Matches 87; Conservative 10; Mismatches 47; Indels 23; Gaps 6;  
 QY 7 ARGP---PAPPEPRSAFSAARTSLVGSADGPGPRHOPLAFDTFVNIGDPDAAAGVFR 64  
 Db 16 ALGPPTPGSSSELSASARTTPLEGTS-----EAAVTFDKYYVNIIGDFDATTQCFR 69  
 QY 65 RLFGAYFFSFTLGKLPKRTLSVTKMKNRDEVQAMIIDGASRR---REMOQSQSVMLALRR 121  
 Db 70 RVGAYFFSFTTACKAKPKNSVTLVNRNDEVQALAFD--KQRRPGARRAASQAMQLDY 127  
 QY 122 GDAVMLLSHDHOGYGAYSNHGKITYTSGFLVYVYDL-----APAP 161  
 Db 128 GDTVMLRLH---GAPHYALGAPGATFSGYLVYADADADAPARGAPAP 171  
 Search completed: January 12, 2004, 08:15:36  
 Job time : 26.6188 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:15:44 ; Search time 20.6986 Seconds  
(without alignments)  
1655.023 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_160\_358

Sequence: 1 ADADAPARPPAPPPERSAR.....LYTPDLAPAPPGAGSELL 170

Scoring table:

BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 747907 seqs, 201509753 residues

Total number of hits satisfying chosen parameters: 747907

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*

2: /cgn2\_6/ptodata/1/pubppaa/PCT\_NEW\_PUB.pep:\*

3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep:\*

4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep:\*

5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep:\*

6: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*

7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep:\*

8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*

9: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*

10: /cgn2\_6/ptodata/1/pubppaa/US09C\_PUBCOMB.pep:\*

11: /cgn2\_6/ptodata/1/pubppaa/US09C\_NEW\_PUB.pep:\*

12: /cgn2\_6/ptodata/1/pubppaa/US09C\_PUBCOMB.pep:\*

13: /cgn2\_6/ptodata/1/pubppaa/US10C\_PUBCOMB.pep:\*

14: /cgn2\_6/ptodata/1/pubppaa/US10C\_PUBCOMB.pep:\*

15: /cgn2\_6/ptodata/1/pubppaa/US10C\_PUBCOMB.pep:\*

16: /cgn2\_6/ptodata/1/pubppaa/US10C\_NEW\_PUB.pep:\*

17: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep:\*

18: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	894	100.0	329	12	US-10-236-055A-14
2	894	100.0	329	12	US-10-085-167-2
3	807.5	90.3	326	12	US-10-236-055A-16
4	237.5	26.6	252	10	US-09-893-737-52
5	230	25.7	202	15	US-10-091-458-43
6	230	25.7	420	12	US-10-236-055A-2
7	230	25.7	746	10	US-09-738-973-185
8	230	25.7	746	10	US-09-854-133-185
9	230	25.7	746	15	US-10-144-649A-185
10	230	25.7	908	12	US-10-309-422-10
11	230	25.7	909	12	US-10-309-422-22
12	230	25.7	957	12	US-10-309-422-14
13	230	25.7	958	12	US-10-309-422-26
14	230	25.7	992	12	US-10-309-422-18
15	230	25.7	993	12	US-10-309-422-30

16	230	25.7	1042	12	US-10-309-422-8	Sequence 8, Appli
17	230	25.7	1043	12	US-10-309-422-20	Sequence 20, Appl
18	230	25.7	1091	12	US-10-309-422-12	Sequence 12, Appl
19	230	25.7	1092	12	US-10-309-422-24	Sequence 24, Appl
20	230	25.7	1126	12	US-10-309-422-16	Sequence 16, Appl
21	230	25.7	1127	12	US-10-309-422-28	Sequence 28, Appl
22	226	25.3	409	12	US-10-236-055A-4	Sequence 4, Appl1
23	220.5	24.7	881	10	US-09-911-176B-44	Sequence 44, Appl1
24	220.5	24.7	881	12	US-10-360-186-44	Sequence 44, Appl1
25	220.5	24.7	881	12	US-10-392-531-44	Sequence 44, Appl1
26	220.5	24.7	881	12	US-10-392-706-44	Sequence 44, Appl1
27	220.5	24.7	881	12	US-10-385-015-5	Sequence 5, Appl1
28	220.5	24.7	881	15	US-10-180-762-44	Sequence 44, Appl1
29	220.5	24.7	881	15	US-10-241-258-44	Sequence 44, Appl1
30	208.5	23.3	198	12	US-10-211-689-16	Sequence 16, Appl1
31	208.5	23.3	207	12	US-10-211-689-10	Sequence 10, Appl1
32	208.5	23.3	207	12	US-10-211-689-12	Sequence 12, Appl1
33	208.5	23.3	214	12	US-10-211-689-14	Sequence 14, Appl1
34	208.5	23.3	243	9	US-09-866-028-42	Sequence 42, Appl1
35	208.5	23.3	243	10	US-09-944-449-42	Sequence 42, Appl1
36	208.5	23.3	243	10	US-09-944-457-42	Sequence 42, Appl1
37	208.5	23.3	243	10	US-09-944-862-42	Sequence 42, Appl1
38	208.5	23.3	243	10	US-09-945-587-42	Sequence 42, Appl1
39	208.5	23.3	243	10	US-09-945-015-42	Sequence 42, Appl1
40	208.5	23.3	243	10	US-09-944-356-42	Sequence 42, Appl1
41	208.5	23.3	243	10	US-09-944-097-42	Sequence 42, Appl1
42	208.5	23.3	243	10	US-09-944-432-42	Sequence 42, Appl1
43	208.5	23.3	243	10	US-09-943-762-42	Sequence 42, Appl1
44	208.5	23.3	243	10	US-09-944-654-42	Sequence 42, Appl1
45	208.5	23.3	243	10	US-09-943-851A-42	Sequence 42, Appl1

ALIGNMENTS

RESULT 1

US-10-236-055A-14

Sequence 14, Application US/10236055A

Publication No. US20030134328A1

GENERAL INFORMATION:

APPLICANT: Basham, Beth E.

APPLICANT: Forsythe, Ian

APPLICANT: Gorman, Daniel M.

APPLICANT: Mateon, Jeanine

APPLICANT: Moshrefi, Mehrdad

APPLICANT: Parham, Christi

TITLE OF INVENTION: MAMMALIAN GENES; RELATED REAGENTS

FILE REFERENCE: DX01343K

CURRENT APPLICATION NUMBER: US/10/236,055A

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/317,988

PRIOR FILING DATE: 2001-09-06

NUMBER OF SEQ ID NOS: 36

SOFTWARE: Patentin version 3.1

SEQ ID NO 14

LENGTH: 329

TYPE: PRT

ORGANISM: Homo sapiens

US-10-236-055A-14

Query Match 100.0%; Score 894; DB 12; Length 329;

Best Local Similarity 100.0%; Pred. No. 7.1e-84;

Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ADADAPARPPAPPPERSARFSAARTRSLVGSAGCPHROPLAFTEFPNIGDFPAAAG 60

160 ADADAPARPPAPPPERSARFSAARTRSLVGSAGCPHROPLAFTEFPNIGDFPAAAG 219

61 VFRCRLPAGYFFSFTLGKLPKRTLSVKLMKNDEVOAMYYDDGASRRRMOSSVLAAR 120

220 VRCRLPAGYFFSFTLGKLPKRTLSVKLMKNDEVOAMYYDDGASRRRMOSSVLAAR 279

121 RDDAVWLASHDHGAGYANNHGKTYTFSGFLVYPLDAPAPPGAGSELL 170

Db 280 RGDVWMLSHDHGAYSNHGKXITFSGLVYPDLAPAAPGLGASELL 329

## RESULT 2

US-10-085-167-2  
; Sequence 2, Application US/10085167  
; Publication No. US20030170781A1  
; GENERAL INFORMATION:  
; APPLICANT: Holloway, James L.  
; APPLICANT: Lok, Si  
; TITLE OF INVENTION: SECRETED PROTEIN ZACRP4  
; FILE REFERENCE: 99-29  
; CURRENT APPLICATION NUMBER: US/10/085.167  
; PRIOR FILING DATE: 2002-02-27  
; PRIOR APPLICATION NUMBER: 60/141,928  
; PRIOR FILING DATE: 1999-07-01  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 329  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-085-167-2

Query Match  
Best Local Similarity 100.0%; Score 894; DB 12; Length 329;  
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ADADAPARGPAPPEPRSAFSAARTSLVSGDAGPGRHOPLAFTTEFVNIGDFDAAAG 60  
Db 160 ADADAPARGPAPPEPRSAFSAARTSLVSGDAGPGRHOPLAFTTEFVNIGDFDAAAG 219  
Qy 61 VFRCLPGAYFFSFTLGKLPKRTLSVKLMKNRDEVQAMTYDDGASRRRMOQSVMALR 120  
Db 220 VFRCLPGAYFFSFTLGKLPKRTLSVKLMKNRDEVQAMTYDDGASRRRMOQSVMALR 279  
Qy 121 RGDVWMLSHDHGAYSNHGKXITFSGLVYPDLAPAAPGLGASELL 170  
Db 280 RGDVWMLSHDHGAYSNHGKXITFSGLVYPDLAPAAPGLGASELL 329

## RESULT 3

US-10-236-055A-16  
; Sequence 16, Application US/10236055A  
; Publication No. US20030134328A1  
; GENERAL INFORMATION:  
; APPLICANT: Basham, Beth E.  
; APPLICANT: Foreythe, Ian  
; APPLICANT: Gorman, Daniel M.  
; APPLICANT: Matteson, Jeanine  
; APPLICANT: Mosnig, Mehdiad  
; APPLICANT: Parham, Christl  
; TITLE OF INVENTION: MAMMALIAN GENES; RELATED REAGENTS  
; FILE REFERENCE: DX01343K  
; CURRENT APPLICATION NUMBER: US/10/236.055A  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/317,988  
; PRIOR FILING DATE: 2001-09-06  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 16  
; LENGTH: 326  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-236-055A-16

Query Match  
Best Local Similarity 90.3%; Score 807.5; DB 12; Length 326;  
Matches 157; Conservative 1; Mismatches 10; Indels 1; Gaps 1;  
Qy 1 ADADAPARGPAPPEPRSAFSAARTSLVSGDAGPGRHOPLAFTTEFVNIGDFDAAAG 60

Db 159 ADADAPARG-PAAPPEPRSAFSAARTSLVSGDAAPGPRHOPLAFTTEFVNIGDFDAAAG 217

Qy 61 VFRCLPGAYFFSFTLGKLPKRTLSVKLMKNRDEVQAMTYDDGASRRRMOQSVMALR 120  
Db 218 VFRCLPGAYFFSFTLGKLPKRTLSVKLMKNRDEVQAMTYDDGASRRRMOQSVMALR 277

Qy 121 RGDVWMLSHDHGAYSNHGKXITFSGLVYPDLAPAAPGLGASEL 169  
Db 278 RGDVWMLSHDHGAYSNHGKXITFSGLVYPDLAPAAPGLGASEL 326

## RESULT 4

US-09-893-737-52  
; Sequence 52, Application US/09893737  
; Patent No. US20020110855A1  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; APPLICANT: Presnell, Scott R.  
; TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS  
; FILE REFERENCE: 00-41  
; CURRENT APPLICATION NUMBER: US/09/893.737  
; PRIOR FILING DATE: 2001-06-28  
; PRIOR APPLICATION NUMBER: US 60/215,446  
; PRIOR FILING DATE: 2000-06-30  
; NUMBER OF SEQ ID NOS: 329  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 52  
; LENGTH: 252  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-893-737-52

Query Match  
Best Local Similarity 43.0%; Score 237.5; DB 10; Length 252;  
Matches 65; Conservative 12; Mismatches 61; Indels 13; Gaps 7;

Qy 9 GPPAPPEPR--SAFSAARTSLVSGDAGPGRHOPLAFTTEFVNIGDFDAAAGVFRCL 66  
Db 106 GPPAPPEPR--SAFSAARTSLVSGDAGPGRHOPLAFTTEFVNIGDFDAAAGVFRCL 160  
Qy 67 PGAYFFSFTLGKLPKRTLSVKLMKNRDEVQAMTYDDGASRRRMOQSVMALR 126  
Db 161 PGAYFFSFTLGKLPKRTLSVKLMKNRDEVQAMTYDDGASRRRMOQSVMALR 218  
Qy 127 --LLSHDHGAYSNHGK--YITFSGLVYP 154  
Db 219 VRMGRDRDN-AITGEGDLYITFSGLVYP 248

## RESULT 5

US-10-091-458-43  
; Sequence 43, Application US/10091458  
; Publication No. US20030068627A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT206C1  
; CURRENT APPLICATION NUMBER: US/10/091.458  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 09/764,900  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/214,886  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 60/217,487  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,758  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,963  
; PRIOR FILING DATE: 2000-07-26

PRIOR APPLICATION NUMBER: 60/217,496  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/225,447  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/218,290  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: 60/225,757  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/226,868  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/216,647  
PRIOR FILING DATE: 2000-07-07  
PRIOR APPLICATION NUMBER: 60/225,267  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/216,880  
PRIOR FILING DATE: 2000-07-07  
PRIOR APPLICATION NUMBER: 60/225,270  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/251,869  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/235,834  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: 60/234,274  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: 60/234,223  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: 60/228,924  
PRIOR FILING DATE: 2000-08-30  
PRIOR APPLICATION NUMBER: 60/224,518  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/236,369  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/224,519  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/220,964  
PRIOR FILING DATE: 2000-07-26  
PRIOR APPLICATION NUMBER: 60/241,809  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/249,299  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/236,327  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/241,785  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/244,617  
PRIOR FILING DATE: 2000-11-01  
PRIOR APPLICATION NUMBER: 60/225,268  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/236,368  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/251,856  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/251,868  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/229,344  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/234,997  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: 60/229,343  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,345  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,287  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,513  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: 60/231,413  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/229,509  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: 60/236,367  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/237,039

PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,038  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/236,370  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/236,802  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,037  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,040  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/240,960  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/239,935  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 60/239,937  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 60/241,787  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/246,474  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/246,532  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/249,216  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,210  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/226,681  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/225,759  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/225,213  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/227,182  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/225,214  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/235,836  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: 60/230,438  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/215,135  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: 60/225,266  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/249,218  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,208  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,213  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,212  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,207  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,245  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,244  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,217  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,211  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,215  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,264  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,214  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,297  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/232,400  
PRIOR FILING DATE: 2000-09-14

```
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08
```

```
Query Match 25.7%; Score 230; DB 15; Length 202;
Best Local Similarity 38.4%; Pred. No. 1.1e-15;
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;
```

```
Oy 4 DAPARGP-----PAPPEPSAFSAARTSLVSGDAGPGRHQPLADTFEYVING 53
Db 49 DVPVTNPATILPVHYVPLPQOMRVAFAAART-----SNLAPGTLDQPIVFDLLNNGE 103
Oy 54 DFDAAAGVRCRLPGAVFSPFTLGKLP-RKTLGVKLMKRNDEVOAMTYDDGASRRREMOS 112
Db 104 TFDLQGRFNCVPNGTYVIFHMLKLVAVNPVLYNLMKNEEVLVSAYANDGAP-DHETAS 162
Oy 113 QSVMLALRGGDAVWMLSHDHGYGA-YSNHGKYITFGSLVYPD 155
Db 163 NHAILQLFQGDQIWLRLH---RGAIYSSWKSTFTSGYLLYOD 202
```

```
RESULT 6
US-10-236-055A-2
; Sequence 2, Application US/10236055A
; Publication No. US20030134328A1
; GENERAL INFORMATION:
; APPLICANT: Basham, Beth E.
; APPLICANT: Foraythe, Ian
; APPLICANT: Gorman, Daniel M.
; APPLICANT: Mattson, Jeanine
; APPLICANT: Moshtrefi, Mehdiad
; APPLICANT: Parham, Christel
; TITLE OF INVENTION: MAMMALIAN GENES; RELATED REAGENTS
; FILE REFERENCE: DX01343K
; CURRENT APPLICATION NUMBER: US/10/236,055A
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/317,988
; PRIOR FILING DATE: 2001-09-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
US-10-236-055A-2
Query Match 25.7%; Score 230; DB 12; Length 420;
Best Local Similarity 38.4%; Pred. No. 2.8e-15;
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;
Oy 4 DAPARGP-----PAPPEPSAFSAARTSLVSGDAGPGRHQPLADTFEYVING 53
Db 267 DVPVTNPATILPVHYVPLPQOMRVAFAAART-----SNLAPGTLDQPIVFDLLNNGE 321
Oy 54 DFDAAAGVRCRLPGAVFSPFTLGKLP-RKTLGVKLMKRNDEVOAMTYDDGASRRREMOS 112
Db 322 TFDLQGRFNCVPNGTYVIFHMLKLVAVNPVLYNLMKNEEVLVSAYANDGAP-DHETAS 380
Oy 113 QSVMLALRGGDAVWMLSHDHGYGA-YSNHGKYITFGSLVYPD 155
Db 381 NHAILQLFQGDQIWLRLH---RGAIYSSWKSTFTSGYLLYOD 420
```

```
RESULT 7
US-09-738-973-185
; Sequence 185, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Flind, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliott, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 185
; LENGTH: 746
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-738-973-185
```

```
Query Match 25.7%; Score 230; DB 10; Length 746;
Best Local Similarity 38.4%; Pred. No. 5.8e-15;
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;
```

```
Oy 4 DAPARGP-----PAPPEPSAFSAARTSLVSGDAGPGRHQPLADTFEYVING 53
Db 593 DVPVTNPATILPVHYVPLPQOMRVAFAAART-----SNLAPGTLDQPIVFDLLNNGE 647
Oy 54 DFDAAAGVRCRLPGAVFSPFTLGKLP-RKTLGVKLMKRNDEVOAMTYDDGASRRREMOS 112
Db 648 TFDLQGRFNCVPNGTYVIFHMLKLVAVNPVLYNLMKNEEVLVSAYANDGAP-DHETAS 706
Oy 113 QSVMLALRGGDAVWMLSHDHGYGA-YSNHGKYITFGSLVYPD 155
Db 707 NHAILQLFQGDQIWLRLH---RGAIYSSWKSTFTSGYLLYOD 746
```

```
RESULT 8
US-09-854-133-185
; Sequence 185, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
```

APPLICANT: Henderson, Robert A.  
APPLICANT: Benson, Darin R.  
APPLICANT: Secrist, Heather  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
FILE REFERENCE: 210121.475C10  
CURRENT APPLICATION NUMBER: US/09/854.133  
NUMBER OF SEQ ID NOS: 735  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 185  
LENGTH: 746  
TYPE: PRF  
ORGANISM: Homo sapien  
US-09-854-133-185

Query Match 25.7%; Score 230; DB 10; Length 746;  
Best Local Similarity 38.4%; Pred. No. 5.8e-15;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 4 DAPARGP-----PAPPEPSAFSAARTSLVSGDAGPGRHOPLAFTFEVNIIG 53  
DB 593 DVPVTNPAATILPVHVYPLPQOMRVAFSAART-----SNLAPGTLDPPIVFDLLNNLIGE 647  
QY 54 DFDAAGVFRCLPGAYFFSFTLGLP-RKTLGVKMKNRDEVQAMTYDDGASRRREMOS 112  
DB 648 TFDLQGRFNCVPNGVTYVFIFHMLKLAIVNPLVYNLMKNEEVLVSAYANDGAP-DHETAS 706  
QY 113 QSWMLALRGGDAVWMLSHDHGYGA-YSNHGKYITTFSGFLVYPP 155  
DB 707 NHAILQLFGDQIWLRLH---RGAIYSSSMKYSTFSGYLLYOD 746

RESULT 9  
US-10-144-649A-185  
Sequence 185, Application US/10144649A  
Publication No. US20030118599A1  
GENERAL INFORMATION:  
APPLICANT: Lodes, Michael J.  
APPLICANT: Wang, Tongtong  
APPLICANT: Fan, Liqun  
APPLICANT: Algate, Paul A.  
APPLICANT: McNeill, Patricia D.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
FILE REFERENCE: 210121.475C11  
CURRENT APPLICATION NUMBER: US/10/144.649A  
NUMBER OF SEQ ID NOS: 749  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 185  
LENGTH: 746  
TYPE: PRF  
ORGANISM: Homo sapien  
US-10-144-649A-185

Query Match 25.7%; Score 230; DB 15; Length 746;  
Best Local Similarity 38.4%; Pred. No. 5.8e-15;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 4 DAPARGP-----PAPPEPSAFSAARTSLVSGDAGPGRHOPLAFTFEVNIIG 53  
DB 593 DVPVTNPAATILPVHVYPLPQOMRVAFSAART-----SNLAPGTLDPPIVFDLLNNLIGE 647  
QY 54 DFDAAGVFRCLPGAYFFSFTLGLP-RKTLGVKMKNRDEVQAMTYDDGASRRREMOS 112  
DB 648 TFDLQGRFNCVPNGVTYVFIFHMLKLAIVNPLVYNLMKNEEVLVSAYANDGAP-DHETAS 706  
QY 113 QSWMLALRGGDAVWMLSHDHGYGA-YSNHGKYITTFSGFLVYPP 155  
DB 707 NHAILQLFGDQIWLRLH---RGAIYSSSMKYSTFSGYLLYOD 746

RESULT 10  
US-10-309-422-10  
Sequence 10, Application US/10309422  
Publication No. US20030139587A1  
GENERAL INFORMATION:  
APPLICANT: Walke, D. Wade  
APPLICANT: Milganowski, Nathaniel L.  
TITLE OF INVENTION: No. US20030139587A1 Human Proteins and Polynucleotides Encodin  
FILE REFERENCE: LEX-0142-USA  
CURRENT APPLICATION NUMBER: US/10/309.422  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10  
LENGTH: 908  
TYPE: PRF  
ORGANISM: homo sapiens  
US-10-309-422-10

Query Match 25.7%; Score 230; DB 12; Length 908;  
Best Local Similarity 38.4%; Pred. No. 7.4e-15;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 4 DAPARGP-----PAPPEPSAFSAARTSLVSGDAGPGRHOPLAFTFEVNIIG 53  
DB 755 DVPVTNPAATILPVHVYPLPQOMRVAFSAART-----SNLAPGTLDPPIVFDLLNNLIGE 809  
QY 54 DFDAAGVFRCLPGAYFFSFTLGLP-RKTLGVKMKNRDEVQAMTYDDGASRRREMOS 112  
DB 810 TFDLQGRFNCVPNGVTYVFIFHMLKLAIVNPLVYNLMKNEEVLVSAYANDGAP-DHETAS 868  
QY 113 QSWMLALRGGDAVWMLSHDHGYGA-YSNHGKYITTFSGFLVYPP 155  
DB 869 NHAILQLFGDQIWLRLH---RGAIYSSSMKYSTFSGYLLYOD 908

RESULT 11  
US-10-309-422-22  
Sequence 22, Application US/10309422  
Publication No. US20030139587A1  
GENERAL INFORMATION:  
APPLICANT: Walke, D. Wade  
APPLICANT: Milganowski, Nathaniel L.  
TITLE OF INVENTION: No. US20030139587A1 Human Proteins and Polynucleotides Encodin  
FILE REFERENCE: LEX-0142-USA  
CURRENT APPLICATION NUMBER: US/10/309.422  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 22  
LENGTH: 909  
TYPE: PRF  
ORGANISM: homo sapiens  
US-10-309-422-22

Query Match 25.7%; Score 230; DB 12; Length 909;  
Best Local Similarity 38.4%; Pred. No. 7.5e-15;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 4 DAPARGP-----PAPPEPSAFSAARTSLVSGDAGPGRHOPLAFTFEVNIIG 53  
DB 756 DVPVTNPAATILPVHVYPLPQOMRVAFSAART-----SNLAPGTLDPPIVFDLLNNLIGE 810

Oy 54 DFDAAAGVRCRLPGAYFFSFTLGKLP-RKTLISVKLMKNRDEVOMIYDDGASRRRENOS 112  
Db 811 TFDLQGRFNCVPVNGTVYVIFPHMLKLVANVPLVYVLMKNRDEVLSAANDGAP-DHETAS 869  
Oy 113 QSVMLALRGDAVWLLSHDHGYGA-YSNHGKYITFSGFLVYPD 155  
Db 870 NHAIIQLFGDQIWLRLH----RGAIVGSSWKYSTFSGYLLYOD 909

## RESULT 12

US-10-309-422-14  
; Sequence 14, Application US/10309422  
; Publication No. US20030139587A1  
; GENERAL INFORMATION:  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wilganowski, Nathaniel L.  
; APPLICANT: Turner, C. Alexander Jr.  
; TITLE OF INVENTION: No. US20030139587A1el Human Proteins and Polynucleotides Encoding  
; FILE REFERENCE: LEX-0142-USA  
; CURRENT APPLICATION NUMBER: US/10/309,422  
; CURRENT FILING DATE: 2002-12-03  
; PRIOR APPLICATION NUMBER: US/09/798,771  
; PRIOR FILING DATE: 2001-03-02  
; PRIOR APPLICATION NUMBER: US 60/186,557  
; PRIOR FILING DATE: 2000-03-02  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-309-422-14

Query Match 25.7%; Score 230; DB 12; Length 957;  
Best Local Similarity 38.4%; Pred. No. 8e-15;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

Oy 4 DAPARGP-----PAPPEPRSAFSAARTSLVSGSDAGPGRHQPPLAFDTFEFVNIIG 53  
Db 804 DVPVTNPAATILPVHVYPLPQOMRVAFSAART-----SNLAPGTLDDQPIVFDLLNNIGE 858  
Oy 54 DFDAAAGVRCRLPGAYFFSFTLGKLP-RKTLISVKLMKNRDEVOMIYDDGASRRRENOS 112  
Db 859 TFDLQGRFNCVPVNGTVYVIFPHMLKLVANVPLVYVLMKNRDEVLSAANDGAP-DHETAS 917  
Oy 113 QSVMLALRGDAVWLLSHDHGYGA-YSNHGKYITFSGFLVYPD 155  
Db 918 NHAIIQLFGDQIWLRLH----RGAIVGSSWKYSTFSGYLLYOD 957

## RESULT 13

US-10-309-422-26  
; Sequence 26, Application US/10309422  
; Publication No. US20030139587A1  
; GENERAL INFORMATION:  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wilganowski, Nathaniel L.  
; APPLICANT: Turner, C. Alexander Jr.  
; TITLE OF INVENTION: No. US20030139587A1el Human Proteins and Polynucleotides Encoding  
; FILE REFERENCE: LEX-0142-USA  
; CURRENT APPLICATION NUMBER: US/10/309,422  
; CURRENT FILING DATE: 2002-12-03  
; PRIOR APPLICATION NUMBER: US/09/798,771  
; PRIOR FILING DATE: 2001-03-02  
; PRIOR APPLICATION NUMBER: US 60/186,557  
; PRIOR FILING DATE: 2000-03-02  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 26  
; LENGTH: 958  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-309-422-26

Query Match 25.7%; Score 230; DB 12; Length 958;  
Best Local Similarity 38.4%; Pred. No. 8e-15;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

Oy 4 DAPARGP-----PAPPEPRSAFSAARTSLVSGSDAGPGRHQPPLAFDTFEFVNIIG 53  
Db 805 DVPVTNPAATILPVHVYPLPQOMRVAFSAART-----SNLAPGTLDDQPIVFDLLNNIGE 859  
Oy 54 DFDAAAGVRCRLPGAYFFSFTLGKLP-RKTLISVKLMKNRDEVOMIYDDGASRRRENOS 112  
Db 860 TFDLQGRFNCVPVNGTVYVIFPHMLKLVANVPLVYVLMKNRDEVLSAANDGAP-DHETAS 918  
Oy 113 QSVMLALRGDAVWLLSHDHGYGA-YSNHGKYITFSGFLVYPD 155  
Db 919 NHAIIQLFGDQIWLRLH----RGAIVGSSWKYSTFSGYLLYOD 958

## RESULT 14

US-10-309-422-18  
; Sequence 18, Application US/10309422  
; Publication No. US20030139587A1  
; GENERAL INFORMATION:  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wilganowski, Nathaniel L.  
; APPLICANT: Turner, C. Alexander Jr.  
; TITLE OF INVENTION: No. US20030139587A1el Human Proteins and Polynucleotides Encoding  
; FILE REFERENCE: LEX-0142-USA  
; CURRENT APPLICATION NUMBER: US/10/309,422  
; CURRENT FILING DATE: 2002-12-03  
; PRIOR APPLICATION NUMBER: US/09/798,771  
; PRIOR FILING DATE: 2001-03-02  
; PRIOR APPLICATION NUMBER: US 60/186,557  
; PRIOR FILING DATE: 2000-03-02  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18  
; LENGTH: 992  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-309-422-18

Query Match 25.7%; Score 230; DB 12; Length 992;  
Best Local Similarity 38.4%; Pred. No. 8.3e-15;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

Oy 4 DAPARGP-----PAPPEPRSAFSAARTSLVSGSDAGPGRHQPPLAFDTFEFVNIIG 53  
Db 839 DVPVTNPAATILPVHVYPLPQOMRVAFSAART-----SNLAPGTLDDQPIVFDLLNNIGE 893  
Oy 54 DFDAAAGVRCRLPGAYFFSFTLGKLP-RKTLISVKLMKNRDEVOMIYDDGASRRRENOS 112  
Db 894 TFDLQGRFNCVPVNGTVYVIFPHMLKLVANVPLVYVLMKNRDEVLSAANDGAP-DHETAS 952  
Oy 113 QSVMLALRGDAVWLLSHDHGYGA-YSNHGKYITFSGFLVYPD 155  
Db 953 NHAIIQLFGDQIWLRLH----RGAIVGSSWKYSTFSGYLLYOD 992

## RESULT 15

US-10-309-422-30  
; Sequence 30, Application US/10309422  
; Publication No. US20030139587A1  
; GENERAL INFORMATION:  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wilganowski, Nathaniel L.  
; APPLICANT: Turner, C. Alexander Jr.  
; TITLE OF INVENTION: No. US20030139587A1el Human Proteins and Polynucleotides Encoding  
; FILE REFERENCE: LEX-0142-USA  
; CURRENT APPLICATION NUMBER: US/10/309,422  
; CURRENT FILING DATE: 2002-12-03  
; PRIOR APPLICATION NUMBER: US/09/798,771  
; PRIOR FILING DATE: 2001-03-02



PRIOR APPLICATION NUMBER: US 60/186,557  
PRIOR FILING DATE: 2000-03-02  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 30  
LENGTH: 993  
TYPE: PRT  
ORGANISM: homo sapiens  
US-10-309-422-30

Query Match 25.7%; Score 230; DB 12; Length 993;  
Best Local Similarity 38.4%; Pred. No. 8.3e-15;  
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 4 DAPARGP-----PAPPRPSAFSAARTSLVSGDAGPGPRRHQPLAFDTFVNIGG 53  
DB 840 DVEPTNPATILPEVHYVPLPQCMRVAFSART-----SNLPGTLDPPIVFDILLNLIGE 894  
QY 54 DFDAAAGVFRCRPLGAYFFSFSLGKLP-RTLSVKLMKRNDEVOAMTYDDGASRRREMOS 112  
DB 895 TFDLQGRFNCVPNGTYVFIHFMLKLAENVPLTYNLMKNEEVLVSAVANDGAP-DHETAS 953  
QY 113 OSYMLALRGGDAVWLLSHDHGYGA-YSNHGKYYTFSGFLVYPD 155  
DB 954 NHAILQLPQGDQIMLRH----RGATYGSSMKYSTFSGYLLYOD 993

Search completed: January 12, 2004, 08:33:37  
Job time : 20.6986 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:14 ; Search time 8.65269 Seconds  
(without alignments)  
831.284 Million cell updates/sec

Title: US-10-085-167-2\_COPY\_160\_358

Sequence: 1 ADADAPARGPAPPPERSAF.....LYTPDLAPAPPLGASSELL 170

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 4231058 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	230	25.7	746	4 US-09-370-838-185	Sequence 185, App
2	220.5	24.7	281	4 US-09-118-408-44	Sequence 44, Appl
3	220.5	24.7	281	4 US-09-506-855-44	Sequence 44, Appl
4	220.5	24.7	281	4 US-09-911-1768-44	Sequence 44, Appl
5	220.5	24.7	281	4 US-09-619-740-44	Sequence 44, Appl
6	220.5	24.7	281	4 US-09-506-852-44	Sequence 44, Appl
7	220.5	23.3	228	4 US-09-336-536-4	Sequence 4, Appl
8	208.5	23.3	243	3 US-09-140-804-2	Sequence 2, Appl
9	208.5	23.3	243	4 US-09-336-536-3	Sequence 3, Appl
10	208.5	23.3	243	4 US-09-686-838B-2	Sequence 2, Appl
11	202.5	22.7	281	3 US-09-118-408-2	Sequence 2, Appl
12	202.5	22.7	281	4 US-09-506-855-2	Sequence 2, Appl
13	202.5	22.7	281	4 US-09-911-1768-2	Sequence 2, Appl
14	202.5	22.7	281	4 US-09-619-740-2	Sequence 2, Appl
15	202.5	22.7	281	4 US-09-506-852-2	Sequence 2, Appl
16	199.5	22.3	243	3 US-09-188-930-295	Sequence 295, App
17	199.5	22.3	243	4 US-09-312-283C-295	Sequence 295, App
18	194.5	21.8	228	4 US-09-336-536-11	Sequence 11, Appl
19	194.5	21.8	228	4 US-09-336-536-10	Sequence 10, Appl
20	194	21.7	259	4 US-09-986-243-47	Sequence 47, Appl
21	194	21.7	260	4 US-09-489-847-198	Sequence 198, App
22	194	21.7	267	4 US-09-489-847-198	Sequence 198, App
23	188	21.0	247	2 US-08-463-911-2	Sequence 2, Appl
24	188	21.0	247	4 US-09-776-976-4	Sequence 4, Appl
25	187.5	21.0	247	4 US-09-909-547-4	Sequence 4, Appl
26	187.5	21.0	294	3 US-09-188-930-294	Sequence 294, App
27	187.5	21.0	294	4 US-09-312-283C-294	Sequence 294, App

28	187	20.9	247	3 US-09-140-804-8	Sequence 8, Appl
29	187	20.9	247	3 US-09-118-408-3	Sequence 3, Appl
30	187	20.9	247	4 US-09-506-855-3	Sequence 3, Appl
31	187	20.9	247	4 US-09-506-855-3	Sequence 3, Appl
32	187	20.9	247	4 US-09-686-838B-8	Sequence 8, Appl
33	187	20.9	247	4 US-09-911-1768-3	Sequence 3, Appl
34	187	20.9	247	4 US-09-619-740-3	Sequence 3, Appl
35	184.5	20.6	285	4 US-09-506-852-3	Sequence 3, Appl
36	182	20.4	247	4 US-09-312-283C-382	Sequence 382, App
37	182	20.4	247	4 US-09-776-976-2	Sequence 2, Appl
38	177.5	19.9	215	3 US-09-140-804-5	Sequence 5, Appl
39	177.5	19.9	215	4 US-09-686-838B-5	Sequence 5, Appl
40	177.5	19.9	215	4 US-09-911-1768-50	Sequence 50, Appl
41	177.5	19.9	215	4 US-09-619-740-53	Sequence 53, Appl
42	175	19.6	231	4 US-09-530-423-2	Sequence 2, Appl
43	175	19.6	244	2 US-08-463-911-7	Sequence 7, Appl
44	175	19.6	244	3 US-09-140-804-3	Sequence 3, Appl
45	175	19.6	244	4 US-09-336-536-20	Sequence 20, Appl

## ALIGNMENTS

```
RESULT 1
US-09-370-838-185
; Sequence 185, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Monahan, Roach
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 185
; LENGTH: 746
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-370-838-185

Query Match      25.7%; Score 230; DB 4; Length 746;
Best Local Similarity 38.4%; Pred. No. 2.6e-17;
Matches 63; Conservative 16; Mismatches 63; Indels 22; Gaps 6;

QY 4 DAPARP-----PAPPEPRSAFSAARTRSVLGSGAPGPRROPLAFDFEFVNIIG 53
DB 593 DPEVNPATIPTEPHVYPLPQQRVAFSART-----SNLAPETLDOPIVFDLLNNLGE 647
QY 54 DEPAAGVFRCLPGAYFFSFTLGLKP-RKTSVVKLMKRDVEYOAMIVDYGASRRREMOS 112
DB 648 TDLQGRNCEVNGVYVIFHMLKLAANVPLVYINMKKEEVLVSAYANDGAP-DHETAS 706
QY 113 QSVMLALRGDAVWLISHDHGIGA-YSNHGKTYTFSGFLVYPD 155
DB 707 NHAIQLFOGDOIWRLH-----RGAIYSSMKYSTFSFGYLIVOD 746

RESULT 2
US-09-118-408-44
; Sequence 44, Application US/09118408A
; Patent No. 6265544
; GENERAL INFORMATION:
; APPLICANT: Shepard, Paul O.
; TITLE OF INVENTION: ADIPOCTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-30
; CURRENT APPLICATION NUMBER: US/09/118,408A
```



Db 259 IFSDEFTYITFSGLVVKPASEP 281

RESULT 6

US-09-506-852-44  
; Sequence 44, Application US/09506852  
; Patent No. 6566499  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS  
; FILE REFERENCE: 97-30  
; CURRENT APPLICATION NUMBER: US/09/506,852  
; CURRENT FILING DATE: 2000-02-17  
; EARLIER APPLICATION NUMBER: 60/053,154  
; EARLIER FILING DATE: 1997-07-18  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 44  
; LENGTH: 281  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-506-852-44

Query Match 24.7%; Score 220.5; DB 4; Length 281;

Best Local Similarity 39.2%; Pred. No. 8,1e-17;  
Matches 56; Conservative 20; Mismatches 56; Indels 9; Gaps 4;

QY 18 SAFAAARTSLVSGDAGPGRHQPPLAFDTEFVNIGDPAAGVRCRLPGAYFFSFTLG 77  
DB 146 AAFSGRKKALHSND-----YFQPVFDTFVNLKHNMTGKRYCYVPGIYFFSLNVH 200  
QY 78 KLPRKTLGVKMKRDEYQAMITDDGASRRREMOSQVMLLRGDVAV--LISHDHDGY 135  
DB 201 TWNCKEYTLHMKKEEV-VILYAO-VSDRSIMOSQSIMBELREDEYVALLFGGERNA 258  
QY 136 GAYSNHGKYTEFSGLVYPDLAP 158  
DB 259 IFSDEFTYITFSGLVVKPASEP 281

RESULT 7

US-09-336-536-4  
; Sequence 4, Application US/09336536  
; Patent No. 6406884  
; GENERAL INFORMATION:  
; APPLICANT: Leidy, K.  
; APPLICANT: McKay, C.  
; APPLICANT: Bossone, S.  
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF  
; FILE REFERENCE: 7853-144  
; CURRENT APPLICATION NUMBER: US/09/336,536  
; CURRENT FILING DATE: 1999-06-18  
; NUMBER OF SEQ ID NOS: 75  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 228  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-336-536-4

Query Match 23.3%; Score 208.5; DB 4; Length 228;

Best Local Similarity 35.7%; Pred. No. 1,4e-15;  
Matches 56; Conservative 18; Mismatches 72; Indels 11; Gaps 3;

QY 4 DAPAGRPAPP-----EPRSAFSAARTSLVSGDAGPGRHQPPLAFDTEFVNIGDPA 58  
DB 70 EAGPAGPTGPAGGCVPRSAFSAKRSERV-----PPSDAPLPFDRVLVNEGHTDAV 124  
QY 59 AGVRCRLPGAYFFSFTLGKLPKRLSVKLMKRNDEVQAMITDDGASRRREMOSQVMLA 118  
DB 125 TGKFTCVPGVYIFA-VHATVYRASLQFDLVKNGESIASFFQFGGMPKPSLSGAMVR 183  
QY 119 LRRGDVAVMLSHDHDGAYSNHGKYTEFSGLVYPDLAP 155

Db 184 LEPEDQVWQVGVDYIGIYASIKTDSITFSGLVYSD 220

RESULT 8

US-09-140-804-2  
; Sequence 2, Application US/09140804  
; Patent No. 6197930  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; APPLICANT: Humes, Jacqueline M.  
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS  
; FILE REFERENCE: 97-49  
; CURRENT APPLICATION NUMBER: US/09/140,804  
; CURRENT FILING DATE: 1998-08-26  
; EARLIER APPLICATION NUMBER: 60/056,983  
; EARLIER FILING DATE: 1997-08-26  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 243  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-140-804-2

Query Match 23.3%; Score 208.5; DB 3; Length 243;

Best Local Similarity 35.7%; Pred. No. 1,5e-15;  
Matches 56; Conservative 18; Mismatches 72; Indels 11; Gaps 3;

QY 4 DAPAGRPAPP-----EPRSAFSAARTSLVSGDAGPGRHQPPLAFDTEFVNIGDPA 58  
DB 85 EAGPAGPTGPAGGCVPRSAFSAKRSERV-----PPSDAPLPFDRVLVNEGHTDAV 139  
QY 59 AGVRCRLPGAYFFSFTLGKLPKRLSVKLMKRNDEVQAMITDDGASRRREMOSQVMLA 118  
DB 140 TGKFTCVPGVYIFA-VHATVYRASLQFDLVKNGESIASFFQFGGMPKPSLSGAMVR 198  
QY 119 LRRGDVAVMLSHDHDGAYSNHGKYTEFSGLVYPDLAP 155  
DB 199 LEPEDQVWQVGVDYIGIYASIKTDSITFSGLVYSD 220

RESULT 9

US-09-336-536-3  
; Sequence 3, Application US/09336536  
; Patent No. 6406884  
; GENERAL INFORMATION:  
; APPLICANT: Leidy, K.  
; APPLICANT: McKay, C.  
; APPLICANT: Bossone, S.  
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF  
; FILE REFERENCE: 7853-144  
; CURRENT APPLICATION NUMBER: US/09/336,536  
; CURRENT FILING DATE: 1999-06-18  
; NUMBER OF SEQ ID NOS: 75  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 243  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-336-536-3

Query Match 23.3%; Score 208.5; DB 4; Length 243;

Best Local Similarity 35.7%; Pred. No. 1,5e-15;  
Matches 56; Conservative 18; Mismatches 72; Indels 11; Gaps 3;

QY 4 DAPAGRPAPP-----EPRSAFSAARTSLVSGDAGPGRHQPPLAFDTEFVNIGDPA 58  
DB 85 EAGPAGPTGPAGGCVPRSAFSAKRSERV-----PPSDAPLPFDRVLVNEGHTDAV 139  
QY 59 AGVRCRLPGAYFFSFTLGKLPKRLSVKLMKRNDEVQAMITDDGASRRREMOSQVMLA 118  
DB 140 TGKFTCVPGVYIFA-VHATVYRASLQFDLVKNGESIASFFQFGGMPKPSLSGAMVR 198

```
QY      119 LRRGDAVWLSDHDDGCAGYNHNKGIITFSGFLVPD 155
        |   |||   |   |||   |
Db      199 LEPEDDQWVVQVGCGVDYIGIVASIKTDSTFSGLVYS D 235
```

```

RESULT 10
US-09-686-838B-2
Sequence 2, Application US/0966838B
Patent No. 6482612
GENERAL INFORMATION:
APPLICANT: Shepard, Paul O.
APPLICANT: Humes, Jacqueline M.
TITLE OF INVENTION: Adipocyte-Specific Protein Homologs
FILE REFERENCE: 97-49D1
CURRENT APPLICATION NUMBER: US/09/666,838B
CURRENT FILING DATE: 2000-10-10
PRIOR APPLICATION NUMBER: US 09/140,804
PRIOR FILING DATE: 1998-08-26
PRIOR APPLICATION NUMBER: US 60/056,983
PRIOR FILING DATE: 1997-08-26
NUMBER OF SEQ ID NOS: 50
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 243
TYPE: PRT
ORGANISM: Homo sapiens
US-09-686-838B-2

```

```

Query Match 23.3% Score 208.5; DB 4 Length 243;
Best local similarity 35.7%; Pred. No. 1.5e-15;
Matches 56; Conservative 18; Mismatches 72; Indels 11; Gaps 3

QY 4 DAPAGPAPP-----EPRSAFSAARTSLVSGSDAGPGRHOPLAFTDEFFVNIQGFDA 58
    |||||
DB 85 EAGPAGPPTGPAGCEGVPPRSATSAKRSERV-----PPSDAPLPFDRLVLTNVEQGHYDAV 139
    |||||

QY 59 AGVPRCLRLGAYFFSEFTLGLPKRTLSVTLMKRDEVQAMITDDGASRRRREMOSQSVMLA 118
    |||||
DB 140 TGEFTTCOVGVYFYFA-VHATVYRASLOPDLVKNGESIASFQFFGGMKPRASISGAMVR 198
    |||||

QY 119 LARGDAVMLSHDHDCYGAYSNHGKITTSGFLVYVD 155
    |||||
DB 199 LEPEDQWVQVGVGDYIGIYASIKTDSITSGFLVSD 235
    |||||

```

```

RESULT 13
US-09-118-408-2
; Sequence 2, Application US/09118408A
; Patent No. 6265544
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-10
; CURRENT APPLICATION NUMBER: US/09/118,408A
; CURRENT FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: 60/053,154
; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-118-408-2

```

```

Query Match 5: Score 203.5; DB 3; Length 281;
      Beef Local Similarity 22.7%;
      Pared. No. 8, 7e-15;
      Matches 60; Conservative 18; Mismatches 58; Indels 33; Gaps 6

QY      5  AARCGPPAD-----PEPR-----SAFSAARTSLVGSADAGPGRH-----QPLAFD 45
DB      119 AARGGTGGKGGKSGMGARCGERCKSHYAAFGVSGRKK-----PMSHSNYHYGTAFD 168

```

Qy 46 TEEVNIIGSPFDAAAGYFRCLRLGAYFPFETLTKPKRLTSLVKMKNDKEVOAMLYDGAS 105  
 Db 169 TEEVNIYDHPNMFQKGYCYVPGILFFPFLNHTNNQKETYLIHKNEEVEVILPAQVG-- 226  
 Qy 106 RREEMSSQSLALRGGDAVWLTHSDHDGAYASNNH--GKVTTPSGFLV 152  
 Db 227 DRRTMSQSLMLELRQDQWVRLYKGERENAIISSELDITYITSGFLV 275

```

RESULT 12
US-09-506-855-2
; Sequence 2, Application US/09506855
; Patent No. 6448221
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Laser, Gerald W.
; TITLE OF INVENTION: INHIBITORS FOR USE IN HEMOSTASIS AND
; TITLE OF INVENTION: IMMUNE FUNCTION
; FILE REFERENCE: 99-12
; CURRENT APPLICATION NUMBER: US/09/506,855
; CURRENT FILING DATE: 2000-02-17
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-506-855-2

```

```

Query Match Similarity      22.7%  Score 202.5; DB 4; Length 281;
Beet Local Similarity      35.5%  Pred No. 8.7e-15;
Match 60; Conservative 18; Mismatches 58; Indels 33; Gaps 6

QY 5 AARGPAP-----PEPR--SAFSAARTSLVSGDAGCPRH---QPLAFD 45
DB 119 AGARHTPKQKQSGMGAPGERCKSHYAFAVSGRAK-----PMSHNYGVQIAD 168

QY 46 TEFVNGDQPDAAACVFCRLPGAYFFSETLCKLPRKTLISVLMKQRDYOAMIDDGAS 105
DB 169 TEFVLYLPHFMNFSTKFCYCVPGTLFFPSLNVHTNMQKETYLTHIMNDEEVLLFAQVG-- 226
QY 106 RRREQSSVMALALRGDAVMLSHDHCYGYASNH--GKYLTTFSGFLV 152
DB 227 DRSINQSSLMETLREQDQVWRVLKGGEEENMIFSEELDTYLTTFSGYLV 275

```

```

RESULT 13
US-09-911-176B-2
/ Sequence 2, Application US/09911176B
/ Patent No. 6518403
/ GENERAL INFORMATION:
/ APPLICANT: Sheppard, Paul O.
/ TITLE OF INVENTION: ANTIBODIES THAT BIND AN
/ TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGY
/ FILE REFERENCE: 97-30D1
/ CURRENT APPLICATION NUMBER: US/09/911,176B
/ CURRENT FILING DATE: 2001-07-23
/ PRIOR APPLICATION NUMBER: 09/118,408
/ PRIOR FILING DATE: 1998-07-17
/ PRIOR APPLICATION NUMBER: 60/053,154
/ PRIOR FILING DATE: 1997-07-18
/ NUMBER OF SEQ ID NOS: 52
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 2
/ LENGTH: 281
/ TYPE: PRT
/ ORGANISM: Homo sapien
US-09-911-176B-2

```

Query Match	22.7%;	Score 202.5;	DB 4;	Length 281;
Best Local Similarity	35.5%;	Pred. No. 8.7e-15;		



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:09 ; Search time 1.79441 Seconds  
(without alignments)  
1661.397 Million cell updates/sec

Title: US-10-085-167-3

Perfect score: 57  
Sequence: 1 FXXXXXXXXXXXXXXXXXXXXXGXFXFX 31Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	38.6	250	2 E83651	hypothetical prote
2	22	38.6	302	2 D82112	tyra protein VC214
3	22	38.6	421	2 A99309	membrane transport
4	22	38.6	433	2 H90490	membrane transport
5	22	38.6	441	2 T19306	hypothetical prote
6	22	38.6	868	2 T02635	D2 protein homolog
7	22	38.6	3705	2 AD0123	probable autotrans
8	22	38.6	4976	2 T14165	peptide synthetase
9	21	36.8	156	2 E72052	rRNA methylase - C
10	21	36.8	156	2 A86573	rRNA methylase [im
11	21	36.8	210	2 F64609	conserved hypochet
12	21	36.8	210	2 B71906	hypothetical prote
13	21	36.8	212	2 H81344	hypothetical prote
14	21	36.8	231	2 B83060	hypothetical prote
15	21	36.8	238	2 T64038	hypothetical prote
16	21	36.8	259	2 T49293	hypothetical prote
17	21	36.8	277	2 T06760	hypothetical prote
18	21	36.8	303	2 T35616	probable membrane
19	21	36.8	303	2 A12543	hypothetical prote
20	21	36.8	307	2 A71602	rifin PRB0955w - m
21	21	36.8	325	2 T31989	hypothetical prote
22	21	36.8	420	2 AG3057	glucose-1-phosphat
23	21	36.8	420	2 G98228	glucose-1-phosphat
24	21	36.8	450	2 B96561	hypothetical prote
25	21	36.8	508	2 T04605	hypothetical prote
26	21	36.8	516	2 C82946	hypothetical prote
27	21	36.8	555	2 T01391	WD-repeat protein
28	21	36.8	653	2 T03319	gene 112 protein -
29	21	36.8	669	2 T24571	hypothetical prote

30	21	36.8	704	2 T13665	NADH2 dehydrogenas
31	21	36.8	713	2 AE0531	lysine decarboxyla
32	21	36.8	724	2 A42953	nitrous oxide redu
33	21	36.8	724	2 D96949	secreted protein c
34	21	36.8	779	2 T44659	nitrous oxide redu
35	21	36.8	826	2 E81706	conserved hypochet
36	21	36.8	885	2 AC2444	hypothetical prote
37	21	36.8	976	2 G82209	GGDEF family prote
38	21	36.8	1138	2 G71554	probable transmemb
39	21	36.8	1316	2 B86240	protein F20B24.12
40	21	36.8	1417	2 H90670	probable adhesin l
41	21	36.8	1417	2 D85521	probable adhesin l
42	20	35.1	40	2 G81731	hypothetical prote
43	20	35.1	119	2 A53257	H+-transporting AT
44	20	35.1	127	2 H64012	sufi protein homol
45	20	35.1	129	2 B90175	NADH dehydrogenase

## ALIGNMENTS

RESULT 1  
E83651  
hypothetical protein BH0013 [imported] - Bacillus halodurans (strain C-125)  
C/Species: Bacillus halodurans  
C/Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #ext\_change 15-Jun-2001  
C/Accession: E83651  
R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiran  
Nucleic Acids Res. 28, 4317-4331, 2000  
A/Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A/Reference number: A83650; WUID:20512582; PMID:11058132  
A/Accession: E83651  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-250 <STO>  
A/Cross-references: GB:AP001507; GB:BA000004; NID:gl0172612; PIDN:BAB03732.1; GSPDB:GN001  
A/Experimental source: strain C-125  
C/Genetics:  
A/Genes: BH0013

Query Match 38.6%; Score 22; DB 2; Length 250;  
Best Local Similarity 36.4%; Pred. No. 2e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXXXXFX 29  
DB 132 FTFSGMSYDF 142

RESULT 2  
D82112  
tyra protein VC2145 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
C/Species: Vibrio cholerae  
C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #ext\_change 02-Feb-2001  
C/Accession: D82112  
R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;  
Chardson, D.; Emolaeva, M.D.; Vamachavan, J.; Baas, S.; Qian, H.; Dragoti, I.; Sellers, P.  
J.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000

A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A/Reference number: A82035; WUID:20406833; PMID:10952301

A/Accession: D82112  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-302 <HEI>

A/Cross-references: GB:AB004287; GB:AB003852; NID:g9656689; PIDN:AAF95290.1; GSPDB:GN001;  
A/Experimental source: serogroup O1; strain N16961; biotype El Tor  
C/Genetics:  
A/Genes: VC2145

A/Map position: 1  
C/Superfamily: Escherichia coli hypothetical protein b2431

Query Match 38.6%; Score 22; DB 2; Length 302;



Query Match 38.6%; Score 22; DB 2; Length 3705;  
Best Local Similarity 36.4%; Pred. No. 3.3e+03;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 3553 FTASVEGTYAF 3563

RESULT 8  
T14165  
peptide synthetase homolog - Mycobacterium smegmatis  
C/Species: Mycobacterium smegmatis  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 01-Dec-2000  
C/Accession: T14165  
R/Yu, S.; Fies, E.; Jacobs Jr., W.R.  
J. Bacteriol. 180, 4676-4685, 1998  
A/Title: Analysis of the exochelin locus in mycobacterium smegmatis: biosynthesis genes  
A/Reference number: Z17898; MUID:98389687; PMID:9721311  
A/Accession: T14165  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-4976 <YUS>  
A/Cross-references: EMBL:AF027770; NID:g3560502; PID:g3560507; PIDN:AAC82550.1  
C/Genetics:  
A/Genes: kDxC  
C/Superfamily: acetate-CoA ligase homology; acyl carrier protein homology  
C/Keywords: carrier protein; phosphopantetheine; phosphoprotein  
F/68-566/Domain: acetate-CoA ligase homology <ACL>  
F/582-649/Domain: acyl carrier protein homology <ACPI>  
F/1620-2062/Domain: acetate-CoA ligase homology <ACPI>  
F/2078-2144/Domain: acyl carrier protein homology <ACPI>  
F/3139-3591/Domain: acetate-CoA ligase homology <ACL2>  
F/3608-3679/Domain: acyl carrier protein homology <ACP3>  
F/4198-4680/Domain: acetate-CoA ligase homology <ACL3>  
F/4637-4705/Domain: acyl carrier protein homology <ACP4>  
F/614.3643,4669/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 38.6%; Score 22; DB 2; Length 4976;  
Best Local Similarity 36.4%; Pred. No. 4.5e+03;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 1738 FAGTSGLYDF 1748

RESULT 9  
E72052  
RNA methylase - Chlamydia pneumoniae (strain CML029)  
C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 03-Nov-2000  
C/Accession: E72052  
R/Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999  
A/Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A/Reference number: A72000; MUID:99206606; PMID:10192388  
A/Accession: E72052  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-156 <ARN>  
A/Cross-references: GB:AE001648; GB:AE001363; NID:g4376946; PIDN:AAD18799.1; PID:g4376995  
A/Experimental source: strain CML029  
C/Genetics:  
A/Genes: spou 2  
C/Superfamily: Chlamydia pneumoniae RNA methylase

Query Match 36.8%; Score 21; DB 2; Length 156;  
Best Local Similarity 36.4%; Pred. No. 2.2e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29

DB 88 FSLPSSGTYVF 98

RESULT 10  
A86573  
RNA methylase (imported) - Chlamydia pneumoniae (strain J138)  
C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C/Accession: A86573  
R/Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise  
Nucleic Acids Res. 28, 2311-2314, 2000  
A/Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.  
A/Reference number: A86491; MUID:20330349; PMID:10871362  
A/Accession: A86573  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-156 <STO>  
A/Cross-references: GB:BA000008; NID:g8579032; PIDN:BA098867.1; GSPDB:GN00142  
A/Experimental source: strain J138  
C/Genetics:  
A/Genes: spou 2  
C/Superfamily: Chlamydia pneumoniae RNA methylase

Query Match 36.8%; Score 21; DB 2; Length 156;  
Best Local Similarity 36.4%; Pred. No. 2.2e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 88 FSLPSSGTYVF 98

RESULT 11  
F64609  
conserved hypothetical integral membrane protein HP0718 - Helicobacter pylori (strain 26)  
C/Species: Helicobacter pylori  
C/Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 17-Mar-2000  
C/Accession: F64609  
R/Tomb, J.F.; White, O.; Kervilave, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;  
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalil, H.G.; Glodek, A.; McKenney  
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weisman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.  
Nature 388, 539-547, 1997  
A/Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser, C.  
A/Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A/Reference number: A64520; MUID:97394467; PMID:9252185  
A/Accession: F64609  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-210 <TOM>  
A/Cross-references: GB:AE000585; GB:AE000511; NID:g2313845; PIDN:AAD07771.1; PID:g2313846  
C/Superfamily: Escherichia coli hypothetical 23k protein (sbm-fba intergenic region)

Query Match 36.8%; Score 21; DB 2; Length 210;  
Best Local Similarity 36.4%; Pred. No. 3e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 75 FGALPTGTYAF 85

RESULT 12  
B71906  
hypothetical protein jhp0656 - Helicobacter pylori (strain J99)  
C/Species: Helicobacter pylori  
A/Variety: strain J99  
C/Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 17-Mar-2000  
C/Accession: B71906  
R/Alm, R.A.; Ling, L.S.; L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;  
Ives, C.; Gibson, R.; Merberg, D.; Miller, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; J  
Nature 397, 176-180, 1999  
A/Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
A/Reference number: A71800; MUID:99120557; PMID:9923682

A:Accession: B71906  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-210 <ARN>  
A:Cross-references: GB:AE001497; GB:AE001439; NID:g4155199; PIDN:AND06232.1; PID:g415520  
A:Experimental source: strain J99  
C:Genetics:  
A:Gene: jhp0656  
C:Superfamily: Escherichia coli hypothetical 23k protein (sbm-fba intergenic region)  
Query Match 36.8%; Score 21; DB 2; Length 210;  
Best Local Similarity 36.4%; Pred. No. 3e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 19 FXXXXXXGYXF 29  
DB 75 FGAVFTGTYAF 85

RESULT 13  
H81344  
hypothetical protein Cj0733 [imported] - Campylobacter jejuni (strain NCTC 11168)  
C:Species: Campylobacter jejuni  
C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
C:Accession: H81344  
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrett  
Nature 403, 665-668, 2000  
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hly  
A:Reference number: A81250; MUID:20150912; PMID:10688204  
A:Accession: H81344  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-212 <PAR>  
A:Cross-references: GB:AL139076; GB:AL111168; NID:g6968128; PIDN:CAB73007.1; PID:g696819  
A:Experimental source: serotype O2, strain NCTC 11168  
C:Genetics:  
A:Gene: Cj0733  
Query Match 36.8%; Score 21; DB 2; Length 212;  
Best Local Similarity 36.4%; Pred. No. 3e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 19 FXXXXXXGYXF 29  
DB 137 FSGKAYGIYSF 147

RESULT 14  
B83060  
hypothetical protein PA4685 [imported] - Pseudomonas aeruginosa (strain PAO1)  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: B83060  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bz  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; Lim,  
J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: B83060  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-231 <STO>  
A:Cross-references: GB:AE004882; GB:AE004091; NID:g9950939; PIDN:AAG08072.1; GSPDB:GN001  
A:Experimental source: strain PAO1  
C:Genetics:  
A:Gene: PA4685

Query Match 36.8%; Score 21; DB 2; Length 231;  
Best Local Similarity 36.4%; Pred. No. 3.3e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 146 FASEDSGVYRF 156

RESULT 15  
I64038  
hypothetical protein H11626 - Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae  
C>Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 08-Oct-1999  
C:Accession: I64038  
R:Feischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirnness, E.F.; Kerlavage, A  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodok, A.; Kelley, J.M.; Weidman, J  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fulhram, J.L.; Geoghegan, N.S.W.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Frazer, C.M.; Smith, H.O.; Venter, C  
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630; PMID:7542800  
A:Accession: I64038  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-238 <TRGR>  
A:Cross-references: GB:U32835; GB:L42023; NID:g1574459; PIDN:AAC23269.1; PID:g1574467; T

Query Match 36.8%; Score 21; DB 2; Length 238;  
Best Local Similarity 36.4%; Pred. No. 3.4e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 122 FALAGMGSTYF 132

Search completed: January 12, 2004, 08:16:52  
Job time : 4.79441 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:09 ; Search time 1.11377 Seconds

(without alignments)  
1308.910 Million cell updates/sec

Title: US-10-085-167-3

Perfect score: 57  
Sequence: 1 FXXXXXXXXXXXXXXXXXXXXXGXXYFX 31

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	38.6	215	YA09_HUMAN	Q9Y3B3 homo sapien
2	22	38.6	409	RUN3_MOUSE	Q64131 m runt-tela
3	21	36.8	175	Y995_THETN	Q8B30 thermoanaer
4	21	36.8	187	Y177_METWA	Q8Q09 methanosarc
5	21	36.8	195	Y614_METAC	Q8Q09 methanosarc
6	21	36.8	196	P24_CRIGR	P49020 cricetulus
7	21	36.8	196	YB05_THRAC	Q9H67 thermoplasm
8	21	36.8	201	P24_HUMAN	Q5363 homo sapien
9	21	36.8	201	P24_MOUSE	Q9R0G3 mus musculu
10	21	36.8	201	P24_RAT	Q63524 ratus norv
11	21	36.8	238	YG26_HAEIN	P42878 haemophilus
12	21	36.8	420	GLGC_AGR75	Q8U815 agrobacteri
13	21	36.8	420	GLGC_AGR75	P39669 agrobacteri
14	21	36.8	420	GLGC_RH1TR	Q9E06 rhizobium t
15	21	36.8	421	GLGC_RH1LO	Q8B534 xanthomonas
16	21	36.8	429	GSA_XANCP	Q8B534 xanthomonas
17	21	36.8	440	MTN5_NOSS7	P35678 nostoc sp.
18	21	36.8	473	L14B_LITER	Q40153 lithospermu
19	21	36.8	724	NOSR_PSEST	Q00780 pseudomonas
20	21	36.8	732	YLH3_SCHPO	Q9H94 schizosaccha
21	21	36.8	738	Y013_BPL2	P42067 naegleria f
22	20	35.1	119	ATP6_NAEFO	P22067 naegleria f
23	20	35.1	127	Y731_HAEIN	P44044 haemophilus
24	20	35.1	173	YGB2_ALCEU	Q44020 alcaligenes
25	20	35.1	175	Y151_FUSNU	Q8R82 fusobacteri
26	20	35.1	181	PAAD_PYRAB	Q9V030 pyrococcus
27	20	35.1	181	PAAD_PYRAB	Q88742 pyrococcus
28	20	35.1	181	PAAD_PYRAB	Q88742 pyrococcus
29	20	35.1	260	GTR5_BOVIN	P58335 bos taurus
30	20	35.1	260	YBXI_BACSU	P44827 bacillus su
31	20	35.1	290	ARY2_MESAU	P50293 mesocricetu
32	20	35.1	290	ARY2_MESAU	P50293 mus musculu
33	20	35.1	290	ARY2_MESAU	P50298 ratus norv

34	20	35.1	297	1	Y060_MYCG	P47306 mycoplasma
35	20	35.1	299	1	Y440_METUA	Q57882 methanococc
36	20	35.1	320	1	GNL_ZYMO	Q01578 zymomonas m
37	20	35.1	334	1	OPST_ORYIA	P87368 oryzias lat
38	20	35.1	336	1	OPST_CARAU	P90309 carassius a
39	20	35.1	368	1	H181_RHIME	Q92860 rhizobium m
40	20	35.1	385	1	Y464_MYCG	P47702 mycoplasma
41	20	35.1	385	1	Y464_MYCPN	P75112 mycoplasma
42	20	35.1	420	1	GLYA_WIGBR	Q8d253 wigleswort
43	20	35.1	421	1	CBP4_HUMAN	Q9u142 homo sapien
44	20	35.1	424	1	GSA_CAMJE	Q9P70 campylobact
45	20	35.1	426	1	GSA_YERPE	Q8b219 yerinia pe

## ALIGNMENTS

RESULT 1	YA09_HUMAN	STANDARD;	PRT;	215 AA.
ID	YA09_HUMAN			
AC	Q9Y3B3;			
DT	15-OCT-2001	(Rel. 40, Created)		
DT	15-OCT-2001	(Rel. 40, Last sequence update)		
DT	15-SEP-2003	(Rel. 42, Last annotation update)		
DE	Hypothetical protein CGI-109 precursor.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
OX	NCBI_Taxid=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20272150; PubMed=10810093;			
RA	Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;			
RT	"Identification of novel human genes evolutionarily conserved in			
RT	Cenorhabditis elegans by comparative proteomics.";			
RL	Genome Res. 10:703-713 (2000).			
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein. Endoplasmic			
CC	reticulum (By similarity).			
CC	-1- SIMILARITY: BELONGS TO THE EMP24/GP25L FAMILY.			
CC	-1- SIMILARITY: Contains 1 GOLD domain.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC				
DR	EMBL; AF151867; AAD34104.1; -			
DR	InterPro; IPR000348; EMP24_GP25L_p24.			
DR	Pfam; PF01105; EMP24_GP25L; 1.			
DR	PROSITE; PS50866; GOLD; 1.			
KW	Hypothetical protein; Transmembrane; Signal;			
KW	Endoplasmic reticulum.			
FT	SIGNAL	1	23	POTENTIAL.
FT	CHAIN	24	215	HYPOTHETICAL PROTEIN CGI-109.
FT	DOMAIN	24	176	LIMENAL (POTENTIAL).
FT	TRANSMEM	177	199	POTENTIAL.
FT	DOMAIN	200	215	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	38	120	GOLD.
SQ	SEQUENCE	215 AA;	24354 MW;	5C4D5BBB8054857B CRC64;
Query Match		38.6%;	Score 22;	DB 1; Length 215;
Best Local Similarity		36.4%;	Pred. No. 74;	
Matches	4;	Conservative	0;	Mismatches 7; Indels 0; Gaps 0;
Qy	19	FXXXXXGXXFX	29	
Db	90	FTASKXGTYKF	100	

RESULT 2

```

RUN3_MOUSE STANDARD; PRT; 409 AA.
ID RUN3_MOUSE
AC Q64131, Q99P92; Q9R199;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Runt-related transcription factor 3 (Core-binding factor, alpha 3 subunit) (CBF-alpha 3) (Acute myeloid leukemia 2 protein) (Oncogene AML-2) (Poliovirus enhancer binding protein 2 alpha C subunit) (PEBP2-alpha C) (PEA2-alpha C) (SL3-3 enhancer factor 1 alpha C subunit) (SL3/ARV core-binding factor alpha C subunit).
DE RuntX3 OR CBF3 OR AML2 OR PEBP2A3.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Spleen;
RA Levanon D., Negreanu V., Groner Y.;
RT "The mouse AML2/Cbf3 cDNA sequence."
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 92-167 FROM N.A.
RX MEDLINE=95331802; PubMed=7607690;
RA Williams C., Speck N.A., Dracopoli N.C., Hofter M.H., Liu P., Collins P.S.;
RT "Identification of a new murine runt domain-containing gene, Cbf3, and localization of the human homolog, CBF3, to chromosome 1p35-pter."
RL Genomics 26:611-614(1995).
RN [3]
RP SEQUENCE OF 1-95 FROM N.A. (ISOFORM 1).
RX MEDLINE=95331801; PubMed=7607689;
RA Calabi F., Rhodes M., Williamson P., Boyd Y.;
RT "Identification and chromosomal mapping of a third mouse runt-like locus."
RL Genomics 26:607-610(1995).
RN [4]
RP SEQUENCE OF 1-95 FROM N.A. (ISOFORM 1).
RC STRAIN=129/Sv;
RA Negreanu V., Levanon D., Bettoun J.D., Groner Y.;
RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.
RN [5]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 2).
RA Bangsow C., Rubins N., Bernstein Y.;
RT "RuntX3 gene structure and function."
RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: BINDS TO THE CORE SITE OF MURINE LEUKEMIA VIRUS, THE CORE SEQUENCES IN THE ENHANCER OF THE POLYOMAVIRUS, AND ALSO TO THE ENHANCERS OF THE T-CELL RECEPTOR GENES. MAY BE INVOLVED IN THE CONTROL OF CELLULAR PROLIFERATION AND/OR DIFFERENTIATION (BY SIMILARITY).
CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. The alpha subunit binds DNA as a monomer and through the runt domain. DNA-binding is increased by heterodimerization. Interacts with TLE1 (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC Name=2;
CC IsoId=Q64131-1; Sequence=Displayed;
CC -1- DOMAIN: A proline/serine/threonine rich region at the C-terminus is necessary for transcriptional activation of target genes.
CC -1- SIMILARITY: Contains 1 Runt domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF155880; AAD38985.1;
DR EMBL; S78518; AAB34843.1;
DR EMBL; AF169246; AAD46381.1;
DR EMBL; AF321443; AAK1181.1;
DR PIR; A56842; A56842.
DR HSSP; O60472; LCMO.
DR MGD; MGI:102672; Runt3.
DR GO; GO:0007411; P:axon guidance; IMP.
DR InterPro; IPR000040; AML1_Runt.
DR Pfam; PF00853; Runt; 1.
DR PRINTS; PR00967; ONCOGENEAML1.
KW Transcription regulation; DNA-binding; Nuclear protein; ATP-binding;
KW Alternative splicing.
FT DOMAIN 23 28 POLY-GLY.
FT NP_BIND 55 182 RUNT.
FT NP_BIND 143 150 ATP (POTENTIAL).
FT VARSPLIC 1 5 R1P1V -> MANSISFSPQLYTNLHT (in isoform 2).
SQ SEQUENCE 409 AA; 43628 MW; 83440C67FC591008 CRC64;

```

```

QY 19 FXXXXXGXYXF 29
DB 331 FYGASGSGYOP 341

```

```

RESULT 3
Y995_THETN STANDARD; PRT; 175 AA.
ID Y995_THETN
AC Q8RB30;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein TTE0995.
GN TTE0995.
OS Thermomicrobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermomicrobacteriales;
OC Thermomicrobacteriaceae; Thermomicrobacter.
CX NCBI_Taxid=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome."
RL Genome Res. 12:689-700(2002).
CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE013064; AAM4250.1;
DR HAMAP; MF_01205; -; 1.
DR InterPro; IPR002589; A1Pp.
DR Pfam; PF01661; A1Pp; 1.
DR SMART; SM00506; A1Pp; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 175 AA; 18762 MW; C1FB4A03A8A1B5A CRC64;

```

```

Query Match          36.8%; Score 21; DB 1; Length 175;
Best Local Similarity 36.4%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy      19 FXXXXXXGXXF 29
      118 FPISTGAYGF 128
Db

RESULT 4
ID Y177 METWA          STANDARD; PRT; 187 AA.
AC 0800F9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypochemical protein MM0177.
GN MM0177.
OS Methanosarcina mazel (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=Geel / G01 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX MEDLINE=22120827; PubMed=1215824;
RA Deppe-Meyer U., Johann A., Hartesch T., Merkl R., Schmitz R.A.,
RA Martinez-Arias R., Henne A., Wieser A., Baumeister S., Jacobl C.,
RA Bruggemann H., Lienard T., Christmann A., Boesmecke M., Steckel S.,
RA Bhattacharyya A., Lykdis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA Fritz H.-J., Gottschalk G.;
RT "The genome of Methanosarcina mazel: evidence for lateral gene
RT transfer between Bacteria and Archaea.";
RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE013241; AAM29873.1; -.
CC DR HAMAP; MF_01205; -. 1.
CC DR InterPro; IPR002589; Alpp.
CC DR Pfam; PF01661; Alpp; 1.
CC KW Hypothetical protein; Complete proteome.
CC SQ SEQUENCE 187 AA; 20361 MW; 8E125281358A1F62 CRC64;

Query Match          36.8%; Score 21; DB 1; Length 187;
Best Local Similarity 36.4%; Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy      19 FXXXXXXGXXF 29
      130 FPISTGAYGF 140
Db

RESULT 5
ID YG14 METAC          STANDARD; PRT; 195 AA.
AC 08TOD0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypochemical protein MA1614.
GN MA1614.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;

```

```

RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nuebaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atwood D., Brown A.,
RA Allen N., Naylor U., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu Y., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umeyan L.A., White O., White R.H., de Macario E.C.,
RA Perry U.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett W.M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE010833; AAM05027.1; -.
CC DR HAMAP; MF_01205; -. 1.
CC DR InterPro; IPR002589; Alpp.
CC DR Pfam; PF01661; Alpp; 1.
CC DR SMART; SM00506; Alpp; 1.
CC KW Hypothetical protein; Complete proteome.
CC SQ SEQUENCE 195 AA; 21671 MW; 35DCEA8ED80C34A0 CRC64;

Query Match          36.8%; Score 21; DB 1; Length 195;
Best Local Similarity 36.4%; Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy      19 FXXXXXXGXXF 29
      138 FPISTGAYGF 148
Db

RESULT 6
ID P24 CRIGR          STANDARD; PRT; 196 AA.
AC P49020;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cop-coated vesicle membrane protein p24 precursor (Fragment).
GN RNP24.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10029;
RN 1;
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=95372409; PubMed=7644530;
RA Stamnes M.A., Craighead M.W., Hoe M.H., Lampen N., Geromano S.,
RA Tempst P., Rothman J.E.;
RT "An integral membrane component of coatamer-coated transport vesicles
RT defines a family of proteins involved in budding.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:8011-8015(1995).
CC -1- FUNCTION: Could have a role in the budding of coatamer-coated and
CC other species of coated vesicles. Could bind cargo molecules to
CC collect them into budding vesicles.
CC -1- SUBUNIT: Interacts with ARFGAP1, inhibiting the GTPase-activating
CC activity of ARFGAP1 (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Golgi-derived

```

CC coatomer-coated vesicles.  
 CC -1- SIMILARITY: BELONGS TO THE EMP24/GP25L FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U26264; AA082925.1; -  
 CC InterPro: IPR000348; EMP24\_gp25L\_p24.  
 CC Pfam: PF01105; EMP24\_GP25L.1.  
 CC PROSITE: PS50866; GOLD.1.  
 CC Transport: Protein transport; Transmembrane; Signal; Golgi stack.  
 CC NON\_TER 1  
 CC SIGNAL <1 15 POTENTIAL.  
 CC CHAIN 16 196 COP-COATED VESICLE MEMBRANE PROTEIN P24.  
 CC DOMAIN 16 163 LUMENAL (POTENTIAL).  
 CC TRANSMEM 164 184 POTENTIAL.  
 CC DOMAIN 185 196 CYTOPLASMIC (POTENTIAL).  
 CC DOMAIN 25 107 GOLD.  
 CC SEQUENCE 196 AA; 22189 MW; 63502CF101B7810 CRC64;  
 SQ  
 Query Match 36.8%; Score 21; DB 1; Length 196;  
 Best Local Similarity 36.4%; Pred. No. 1.2e+02;  
 Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 19 FXXXXXXGYXF 29  
 Db 77 FAAMHWDGYKXF 87  
 RESULT 7  
 YB05 THEAC STANDARD; PRT; 196 AA.  
 ID YB05 THEAC  
 AC Q9HJ67;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein Ta1105.  
 GN Ta1105.  
 OS Thermoplasma acidophilum.  
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;  
 OC Thermoplasmataceae; Thermoplasma.  
 CC NCBI\_TaxID=2303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM 1728;  
 RX MEDLINE=20479972; PubMed=11029001;  
 RA Ruepp A., Grail W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
 RA Mewes H.-W., Fishman D., Stocker S., Lupas A.N., Baumeister W.;  
 RT "The genome sequence of the thermophilic scavenger Thermoplasma  
 RT acidophilum."  
 RL Nature 407:508-513(2000).  
 CC -1- SIMILARITY: BELONGS TO THE UPR0189 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AL445066; CAC12232.1; ALT\_INIT.  
 CC HAMAP: MF\_01205; -1.  
 CC InterPro: IPR002589; A1pp.  
 CC Pfam: PF01661; A1pp.1.  
 CC SMART: SM00506; A1pp.1.  
 KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 196 AA; 21664 MW; 06F0D91D7665E2 CRC64;  
 Query Match 36.8%; Score 21; DB 1; Length 196;  
 Best Local Similarity 36.4%; Pred. No. 1.2e+02;  
 Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 19 FXXXXXXGYXF 29  
 Db 126 FPA1STGAYGF 136  
 RESULT 8  
 P24 HUMAN STANDARD; PRT; 201 AA.  
 ID P24 HUMAN  
 AC Q15363;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE COP-coated vesicle membrane protein p24 precursor (p24A).  
 GN RNP24.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 CC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain.  
 RX MEDLINE=96291865; PubMed=8663407;  
 RA Blum R., Felck P., Puyde M., Vandekerckhove J., Klengel R.,  
 RA Naetanczyk W., Schulz I.;  
 RT "Imp21 and p24A, two type I proteins enriched in pancreatic  
 RT microsomal membranes, are members of a protein family involved in  
 RT vesicular trafficking."  
 RL J. Biol. Chem. 271:17183-17189(1996).  
 CC -1- FUNCTION: Could have a role in the budding of coatomer-coated and  
 CC other species of coated vesicles. Could bind cargo molecules to  
 CC collect them into budding vesicles.  
 CC -1- SUBUNIT: Interacts with ARFGAP1, inhibiting the GTPase-activating  
 CC activity of ARFGAP1 (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Golgi-derived  
 CC coatomer-coated vesicles.  
 CC -1- SIMILARITY: BELONGS TO THE EMP24/GP25L FAMILY.  
 CC -1- SIMILARITY: Contains 1 GOLD domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X92098; CAA63069.1; -  
 CC GO: GO:0016021; C:Integral to membrane; TAS.  
 CC GO: GO:0005792; C:mitosome; TAS.  
 CC GO: GO:0005886; C:plasma membrane; TAS.  
 CC GO: GO:0006886; P:intracellular protein transport; TAS.  
 CC InterPro: IPR000348; EMP24\_gp25L\_p24.  
 CC Pfam: PF01105; EMP24\_GP25L.1.  
 CC PROSITE: PS50866; GOLD.1.  
 CC Transport: Protein transport; Transmembrane; Signal; Golgi stack.  
 CC NON\_TER 1  
 CC SIGNAL 20  
 CC CHAIN 21 201 COP-COATED VESICLE MEMBRANE PROTEIN P24.  
 CC DOMAIN 21 168 LUMENAL (POTENTIAL).  
 CC TRANSMEM 169 189 POTENTIAL.  
 CC DOMAIN 190 201 CYTOPLASMIC (POTENTIAL).  
 CC DOMAIN 30 112 GOLD.  
 CC SEQUENCE 201 AA; 22761 MW; C452370E459DC894 CRC64;  
 SQ  
 Query Match 36.8%; Score 21; DB 1; Length 201;  
 Best Local Similarity 36.4%; Pred. No. 1.3e+02;  
 Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;



QY 19 FXXXXXXGXXF 29  
DB 82 FAAMHDTYKF 92

## RESULT 9

ID\_P24\_MOUSE STANDARD; PRT; 201 AA.  
AC Q9R0Q3;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Cop-coated vesicle membrane protein p24 precursor (p24A).  
GN RNP24 OR SID394.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;

RP SEQUENCE FROM N.A.  
RA Seki N., Hattori A., Hayaashi A., Kozuma S., Muramatsu M., Saito T.;  
RT "Mouse mRNA for transmembrane protein."  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: Could have a role in the budding of coatmer-coated and  
CC other species of coated vesicles. Could bind cargo molecules to  
CC collect them into budding vesicles (By similarity).  
CC -1- SUBUNIT: Interacts with ARFGAP1, inhibiting the GTPase-activating  
CC activity of ARFGAP1 (By similarity).  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Golgi-derived  
CC coatmer-coated vesicles (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE EMP24/GP25L FAMILY.  
CC -1- SIMILARITY: Contains 1 GOLD domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AB025218; BAA84689.1; -  
DR MGD: MGI:1929269; Rnp24.  
DR InterPro: IPR000348; Emp24\_GP25L\_p24.  
DR Pfam: PF01105; Emp24\_GP25L; 1.  
DR PROSITE: PS50866; GOLD; 1.  
KM Transport; Protein transport; Transmembrane; Signal; Golgi stack.  
FT CHAIN 1 20 POTENTIAL.  
FT DOMAIN 21 201 COP-COATED VESICLE MEMBRANE PROTEIN P24.  
FT TRANSMEM 169 189 POTENTIAL.  
FT DOMAIN 190 201 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 30 112 GOLD.  
SQ SEQUENCE 201 AA; 22705 MW; F5DE259FC419CF85 CRC64;

Query Match 36.8%; Score 21; DB 1; Length 201;  
Best Local Similarity 36.4%; Pred. No. 1.3e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGXXF 29  
DB 82 FAAMHDTYKF 92

## RESULT 10

ID\_P24\_RAT STANDARD; PRT; 201 AA.  
AC Q63524;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Cop-coated vesicle membrane protein p24 precursor (p24A) (RNP21.4).  
GN RNP24.

OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;

RP SEQUENCE FROM N.A.  
RA STRAIN=Wistar; TISSUE=Pancreas;  
RX MEDLINE=96291865; PubMed=8663407;  
RA Blum R., Feick P., Puyse P., Vandekerckhove J., Klengel R.,  
RA Nasalencyk W., Schulz I.;  
RT "mp21 and p24, two type I proteins enriched in pancreatic  
RT microsomal membranes, are members of a protein family involved in  
RT vesicular trafficking.";  
RL J. Biol. Chem. 271:17183-17189 (1996).

RP INTERACTION WITH ARFGAP1.  
RX MEDLINE=21629435; PubMed=11748249;  
RA Lanoix J., Ouwendijk J., Stark A., Szafer E., Cassel D., Dejgaard K.,  
RA Weiss M., Nilsson T.;  
RT "Sorting of Golgi resident proteins into different subpopulations of  
RT COP1 vesicles: a role for Arfgap1.";  
RL J. Cell Biol. 155:1199-1212 (2001).  
CC -1- FUNCTION: Could have a role in the budding of coatmer-coated and  
CC other species of coated vesicles. Could bind cargo molecules to  
CC collect them into budding vesicles.  
CC -1- SUBUNIT: Interacts with ARFGAP1, inhibiting the GTPase-activating  
CC activity of ARFGAP1.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Golgi-derived  
CC coatmer-coated vesicles.  
CC -1- SIMILARITY: BELONGS TO THE EMP24/GP25L FAMILY.  
CC -1- SIMILARITY: Contains 1 GOLD domain.  
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X92097; CAA63068.1; -  
DR InterPro: IPR000348; Emp24\_GP25L\_p24.  
DR Pfam: PF01105; Emp24\_GP25L; 1.  
DR PROSITE: PS50866; GOLD; 1.  
KM Transport; Protein transport; Transmembrane; Signal; Golgi stack.  
FT CHAIN 1 20 POTENTIAL.  
FT DOMAIN 21 201 COP-COATED VESICLE MEMBRANE PROTEIN P24.  
FT TRANSMEM 169 189 POTENTIAL.  
FT DOMAIN 190 201 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 30 112 GOLD.  
SQ SEQUENCE 201 AA; 22733 MW; E35A3816429DDF9A CRC64;

Query Match 36.8%; Score 21; DB 1; Length 201;  
Best Local Similarity 36.4%; Pred. No. 1.3e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGXXF 29  
DB 82 FAAMHDTYKF 92

## RESULT 11

ID\_YG26\_HAEIN STANDARD; PRT; 238 AA.  
AC P44278;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein H11626.  
GN H11626.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

```

OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RD / KM20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Uettermann J.F., Phillips R., Spriggs T., Hedblom B., Corton M.D.,
RA Uettermann T.R., Hanna M.C., Nguyen D.T., Saubol D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Frazer C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: TO B.SUBTILIS YWIC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U3835; AAC23269.1; -.
DR PIR: I64038; I64038.
DR TIGR: H11626; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 19 39 POTENTIAL.
FT TRANSMEM 64 84 POTENTIAL.
FT TRANSMEM 85 105 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 141 161 POTENTIAL.
FT TRANSMEM 176 196 POTENTIAL.
FT TRANSMEM 218 238 POTENTIAL.
SO SEQUENCE 238 AA; 27573 MW; 48B2034A9F657DA CRC64;

Query Match 36.8%; Score 21; DB 1; Length 238;
Best Local Similarity 36.4%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 19 FXXXXXXGXXF 29
Db 122 PALMGSGTYF 132

RESULT 12
GLGC_AGR5
ID GLGC_AGR5 STANDARD; PRT; 420 AA.
AC Q8U8L5;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glucose-1-phosphate adenylyltransferase (EC 2.7.7.27) (ADP-glucose
DE synthase) (ADP-glucose pyrophosphorylase) (ADPGlc Phase).
GN GLGC OR ATU4076 OR AGR_L_1560.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=16299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Stubbins J.C., Kaul R., Monks D.E., Kitejima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F.Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boye D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,

```

```

RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M., Mullin L.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Hummel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wolanin C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
CC -1- CATALYTIC ACTIVITY: ATP + alpha-D-glucose 1-phosphate =
CC dihydroxyacetone phosphate + ADP-glucose.
CC -1- PATHWAY: Glycogen biosynthesis; first step.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL AND PLANTS GLUCOSE-1-
CC PHOSPHATE ADENYLYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB009376; AAL4487.1; -.
DR EMBL: AB008276; AAK69353.1; -.
DR PIR: AG3057; AG3057.
DR PIR: G98228; G98228.
DR HAMAP: MF 00624; -.
DR InterPro: IPR005836; ADP_Glu_pyroph.
DR InterPro: IPR005835; NTP_transferase.
DR Pfam: PF00483; NTP_transferase_1.
DR PROSITE: PS00808; ADP_GLC_PYROPHOSPH_1; 1.
DR PROSITE: PS00809; ADP_GLC_PYROPHOSPH_2; 1.
DR PROSITE: PS00810; ADP_GLC_PYROPHOSPH_3; 1.
KW Glycogen biosynthesis; Transferase; Nucleotidyltransferase;
KW Complete proteome.
SO SEQUENCE 420 AA; 47043 MW; 6F6BC076EFC27484 CRC64;

Query Match 36.8%; Score 21; DB 1; Length 420;
Best Local Similarity 36.4%; Pred. No. 2.5e+02;
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 19 FXXXXXXGXXF 29
Db 201 PALMGSGTYF 211

RESULT 13
GLGC_AGR5
ID GLGC_AGR5 STANDARD; PRT; 420 AA.
AC P39669;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glucose-1-phosphate adenylyltransferase (EC 2.7.7.27) (ADP-glucose
DE synthase) (ADP-glucose pyrophosphorylase) (ADPGlc Phase).
GN GLGC.
OS Agrobacterium tumefaciens.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=358;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=AA348;
RX MEDLINE=95047522; PubMed=7959036;
RA Utiato A.D., Ugalde R.A.;
RT "A chromosomal cluster of genes encoding ADP-glucose synthetase,
RT glucose synthase and phosphoglucosyltransferase in Agrobacterium
RT tumefaciens";
RL Gene 150:117-122(1994).
CC -1- CATALYTIC ACTIVITY: ATP + alpha-D-glucose 1-phosphate =
CC diophosphate + ADP-glucose.
CC -1- PATHWAY: Glycogen biosynthesis; first step.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL AND PLANTS GLUCOSE-1-
CC PHOSPHATE ADENYLYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR HAMAP: MF_00624; -; 1.
DR InterPro: IPR005835; ADP_Glu_pyrop.
DR InterPro: IPR005835; NTP_transferase.
DR Pfam: PF00483; NTP_transferase; 1.
DR PROSITE: PS00808; ADP_GLC_PYROPHOSPH_1; 1.
DR PROSITE: PS00809; ADP_GLC_PYROPHOSPH_2; 1.
DR PROSITE: PS00810; ADP_GLC_PYROPHOSPH_3; 1.
KW Glycogen biosynthesis; Transferase; Nucleotidyltransferase.
SQ SEQUENCE 420 AA; 47029 MW; F93F5E4B596E698 CRC64;

Query Match 36.8%; Score 21; DB 1; Length 420;
Best Local Similarity 36.4%; Pred. No. 2.5e+02;
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 19 FXXXXXXGXXF 29
Db 201 FALASMGIVF 211

RESULT 14
GLGC_RHITR STANDARD; PRT; 420 AA.
AC 09EUT6;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glucose-1-phosphate adenylyltransferase (EC 2.7.7.27) (ADP-glucose
DE synthase) (ADP-glucose pyrophosphorylase) (ADPGlc Pphase).
GN GLGC.
OS Rhizobium tropici.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=398;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CIAT899;
RX MEDLINE=21142510; PubMed=11208782;
RA Maricouli S., Zorreguieta A., Santamaria C., Temprano F., Sberon M.,
RA Megias M., Downie J.A.;
RT "Enhanced symbiotic performance by Rhizobium tropici glycogen synthase
RT mutants";
RL Bacteriol. 183:854-864(2001).
CC -1- CATALYTIC ACTIVITY: ATP + alpha-D-glucose 1-phosphate =
CC diophosphate + ADP-glucose.
CC -1- PATHWAY: Glycogen biosynthesis; first step.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL AND PLANTS GLUCOSE-1-
CC PHOSPHATE ADENYLYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR HAMAP: MF_00624; CAC17471.1; -;
DR InterPro: IPR005835; ADP_Glu_pyrop.
DR InterPro: IPR005835; NTP_transferase.
DR Pfam: PF00483; NTP_transferase; 1.
DR PROSITE: PS00808; ADP_GLC_PYROPHOSPH_1; 1.
DR PROSITE: PS00809; ADP_GLC_PYROPHOSPH_2; 1.
DR PROSITE: PS00810; ADP_GLC_PYROPHOSPH_3; 1.
KW Glycogen biosynthesis; Transferase; Nucleotidyltransferase.
SQ SEQUENCE 420 AA; 47006 MW; 491D3C73EE80767 CRC64;

Query Match 36.8%; Score 21; DB 1; Length 420;
Best Local Similarity 36.4%; Pred. No. 2.5e+02;
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 19 FXXXXXXGXXF 29
Db 201 FALASMGIVF 211

RESULT 15
GLGC_RHITR STANDARD; PRT; 421 AA.
AC 0985F3; 0930E7;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glucose-1-phosphate adenylyltransferase (EC 2.7.7.27) (ADP-glucose
DE synthase) (ADP-glucose pyrophosphorylase) (ADPGlc Pphase).
GN GLGC OR MLR7588.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RA Lepek V.C., Tomatis P.E., Giambiagi S., Ugalde R.A.;
RT "Partial characterization of R. loti gene cluster for glycogen
RT metabolism: differences with A. tumefaciens";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Matanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsunoto M., Matsuno A.,
RA Nouchizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
CC -1- CATALYTIC ACTIVITY: ATP + alpha-D-glucose 1-phosphate =
CC diophosphate + ADP-glucose.
CC -1- PATHWAY: Glycogen biosynthesis; first step.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL AND PLANTS GLUCOSE-1-
CC PHOSPHATE ADENYLYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR HAMAP: MF_00624; AAK58595.1; -;

```

DR EMBL; AP003012; BAB54019.1; -.  
 DR HAMAP; MF 00624; -; 1.  
 DR InterPro; IPR005836; ADP\_Glu\_Pyrop.  
 DR InterPro; IPR005835; NTP\_transferase.  
 DR Pfam; PF00483; NTP\_transferase; 1.  
 DR PROSITE; PS00808; ADP\_GLC\_PYROPHOSPH\_1; 1.  
 DR PROSITE; PS00809; ADP\_GLC\_PYROPHOSPH\_2; 1.  
 DR PROSITE; PS00810; ADP\_GLC\_PYROPHOSPH\_3; 1.  
 KM Glycogen biosynthesis; Transferase; Nucleotidyltransferase;  
 KM Complete proteome.  
 FT CONFLICT 145 145 M -> L (IN REF. 1).  
 FT CONFLICT 154 154 N -> G (IN REF. 1).  
 FT CONFLICT 201 201 D -> E (IN REF. 1).  
 FT CONFLICT 227 227 E -> D (IN REF. 1).  
 FT CONFLICT 262 262 A -> M (IN REF. 1).  
 FT CONFLICT 287 287 I -> V (IN REF. 1).  
 FT CONFLICT 338 338 T -> S (IN REF. 1).  
 FT CONFLICT 355 355 T -> A (IN REF. 1).  
 FT CONFLICT 395 395 A -> V (IN REF. 1).  
 FT CONFLICT 411 411 I -> V (IN REF. 1).  
 SQ SEQUENCE 421 AA; 47036 MM; 6DD17FC327A876D9 CRC64;

Query March 36.8%; Score 21; DB 1; Length 421;  
 Best Local Similarity 36.4%; Pred. No. 2.Se+02;  
 Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 19 FXXXXGXYP 29  
 Db 202 PALASMGIVP 212

Search completed: January 12, 2004, 08:13:04  
 Job time : 4.11377 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:10 ; Search time 4.26946 Seconds

(without alignments)  
1873.686 Million cell updates/sec

Title: US-10-085-167-3

Perfect score: 57  
Sequence: 1 FXXXXXXXXXXXXXXXXXXXXXXXGXFXFX 31

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP archaea: \*  
2: SP bacteria: \*  
3: SP fungi: \*  
4: SP human: \*  
5: SP invertebrate: \*  
6: SP mammal: \*  
7: SP mhc: \*  
8: SP organelle: \*  
9: SP phage: \*  
10: SP plant: \*  
11: SP rodent: \*  
12: SP virus: \*  
13: SP vertebrate: \*  
14: SP unclassified: \*  
15: SP rivirus: \*  
16: SP bacteriaph: \*  
17: SP archaeop: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	40.4	352	13	Q9DED3
2	23	40.4	618	16	Q921N9
3	22	38.6	144	11	Q8BRJ1
4	22	38.6	188	4	Q8NBD8
5	22	38.6	224	4	Q8WU6
6	22	38.6	224	4	Q96K51
7	22	38.6	250	16	Q9KGP4
8	22	38.6	290	16	Q8CS05
9	22	38.6	296	16	Q9CNR6
10	22	38.6	300	16	Q8DBA0
11	22	38.6	302	16	Q9XQ59
12	22	38.6	333	17	Q8PXN5
13	22	38.6	335	17	Q8PSM8
14	22	38.6	355	5	Q966G5
15	22	38.6	384	5	Q8EJZ4
16	22	38.6	404	17	Q8U1U7

17	22	38.6	409	11	Q91ZK1	Q91xk1 rattus norv
18	22	38.6	421	17	Q97Y43	Q97y43 sulfolobus
19	22	38.6	423	17	Q97UD8	Q97ud8 sulfolobus
20	22	38.6	440	16	Q8XPA4	Q8xpa4 clostridium
21	22	38.6	441	5	Q18017	Q18017 caenorhabdi
22	22	38.6	529	10	Q9RCF7	Q9rcf7 arabidopsis
23	22	38.6	679	16	Q8XW21	Q8xw21 ralslonia s
24	22	38.6	787	5	Q9GSD4	Q9gsd4 plasmodium
25	22	38.6	868	5	Q60984	Q60984 dictyosteli
26	22	38.6	1127	2	Q81J44	Q81j44 bacillus st
27	22	38.6	1457	2	Q9F289	Q9f289 yersinia pe
28	22	38.6	1677	10	Q9AS94	Q9as94 oryza sativ
29	22	38.6	1913	5	Q9GR96	Q9gr96 leucophaea
30	22	38.6	3705	2	Q9F285	Q9f285 yersinia pe
31	22	38.6	3705	16	Q8ZHA1	Q8zha1 yersinia pe
32	22	38.6	3710	16	Q8CZU2	Q8czu2 yersinia pe
33	22	38.6	4976	2	Q87J14	Q87j14 mycobacteri
34	22	38.6	5020	16	Q8E9W3	Q8e9w3 shewanella
35	22	38.6	45	16	Q8F630	Q8f630 leptospira
36	21	36.8	109	5	Q9NHR7	Q9nh7 plasmodium
37	21	36.8	109	5	Q9NHL5	Q9nh5 plasmodium
38	21	36.8	113	5	Q8MNR3	Q8mnr3 dictyosteli
39	21	36.8	135	16	Q8PAA7	Q8pa7 escherichia
40	21	36.8	154	2	Q9LFR0	Q9lfr0 salmonella
41	21	36.8	154	16	Q9J5R2	Q9j5r2 salmonella
42	21	36.8	156	16	Q9Z7P4	Q9z7p4 chlamydia p
43	21	36.8	159	11	Q8C1Z9	Q8c1z9 mus musculu
44	21	36.8	169	16	Q8E5W2	Q8e5w2 streptococc
45	21	36.8	169	16	Q8E085	Q8e085 streptococc

## ALIGNMENTS

### RESULT 1

ID Q9DED3 PRELIMINARY; PRT; 352 AA.  
AC Q9DED3;  
DT 01-MAR-2001 (T-EMBLrel. 16, Created)  
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)  
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)  
DE Inhibin.  
GN INH.  
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
OC Proteocephalestygii; Salmoniformes; Salmonidae; Oncorhynchus.  
OX NCBI\_TaxID=8022;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Tada T., Endo M., Hirono I., Takashima F., Aoki T.;  
RT "Differential expression and cellular localization of activin and  
inhibin mRNA in the rainbow trout ovary and testis.";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBD databases.  
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.  
DR EMBL; AB044566; BAB19272.1; -  
DR InterPro; IPR002405; Inhibin\_alpha.  
DR InterPro; IPR001839; TGFp.  
DR Pfam; PF00019; TGF-beta; 1.  
DR PRINTS; PR00669; INHIBINA.  
DR PRODOM; PD000357; TGFp; 1.  
DR SMART; SM00204; TGFp; 1.  
DR PROSITE; PS00250; TGF-BETA\_1; 1.  
SQ SEQUENCE 352 AA; 3571 MW; ED9CCE860F912ED CRC64;

Query Match 40.4%; Score 23; DB 13; Length 352;  
Best local Similarity 36.4%; Pred. No. 5.1e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGXFX 29  
DB 326 FTTSDGYSYF 336

```

RESULT 2
ID Q92LN9 PRELIMINARY; PRT; 618 AA.
AC Q92LN9
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein R02996.
GN R02996 OR SMC03097.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Goffie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Maury D.,
RA Poll T., Portetelle D., Puehler A., Purnelle B., Rampeyger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.,
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591792; CAC47575.1; -
DR InterPro; IPR001084; Bac_surfAg_D15.
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR001412; CrnA-synt_1.
DR Pfam; PF01103; Bac_surfAg_1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 618 AA; 65656 MW; E51276B29B1FB78 CRC64;

Query Match 40.4%; Score 23; DB 16; Length 618;
Best Local Similarity 36.4%; Pred. No. 9.1e+02;
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXGXYXF 29
Db 464 FEASAGSYAF 474

RESULT 3
ID Q8BPJ1 PRELIMINARY; PRT; 144 AA.
AC Q8BPJ1
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Runt related transcription factor 3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eye;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA The FANTOM Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK053910; BAC35587.1; -
SQ SEQUENCE 144 AA; 14731 MW; EC4789698ADB9D7 CRC64;

Query Match 38.6%; Score 22; DB 11; Length 144;
Best Local Similarity 36.4%; Pred. No. 3.7e+02;
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

QY 19 FXXXXXGXYXF 29
Db 66 FYGASGSYQF 76

RESULT 4
ID Q8NB08 PRELIMINARY; PRT; 188 AA.
AC Q8NB08
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein FLJ90737.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX Isoqat T., Oca T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Isoqat T., Oca T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto Y., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA Masuno Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RT "NEDO human cDNA sequencing project."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK075218; BAC11479.1; -
DR InterPro; IPR000348; Emp24_gp25L_p24.
DR Pfam; PF01105; Emp24_gp25L; 1.
KM Hypothetical protein.
SQ SEQUENCE 188 AA; 21233 MW; 15A10CC4BD553DB3 CRC64;

Query Match 38.6%; Score 22; DB 4; Length 188;
Best Local Similarity 36.4%; Pred. No. 4.9e+02;
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXGXYXF 29
Db 98 FTASKNGTYKF 108

RESULT 5
ID Q8WU06 PRELIMINARY; PRT; 224 AA.
AC Q8WU06
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical protein FLJ90481.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC Isoqat T., Oca T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Isoqat T., Oca T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto Y., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA Masuno Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RT "NEDO human cDNA sequencing project."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC019349; AAH19349.1; -
DR EMBL; AK074962; BAC11318.1; -
DR InterPro; IPR000348; Emp24_gp25L_p24.
DR Pfam; PF01105; Emp24_gp25L; 1.
KM Hypothetical protein.

```

SQ SEQUENCE 224 AA; 25171 MW; 9D0D2110579837B3 CRC64;  
 Query Match 38.6%; Score 22; DB 4; Length 224;  
 Best Local Similarity 36.4%; Pred. No. 5.8e+02;  
 Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 19 FXXXXXXGYXF 29  
 DB 98 FTASKNGTYKF 108  
 RESULT 6  
 Q96K51 PRELIMINARY; PRT; 224 AA.  
 AC Q96K51;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Hypothetical protein FLJ14606.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Shiraori A., Sudo H.,  
 RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
 RA Yamanabe S., Kimura K., Murakami K., Ichii S., Kawai Y., Saito K.,  
 RA Nimoto K., Iwayanagi T.;  
 RT "NEO human cDNA sequencing project."  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK027512; BAB5166.1; -;  
 DR InterPro: IPR000348; EMP24\_GP25L\_p24.  
 DR Pfam: PF01105; EMP24\_GP25L; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 224 AA; 25181 MW; 9D0D2113989837B3 CRC64;  
 Query Match 38.6%; Score 22; DB 4; Length 224;  
 Best Local Similarity 36.4%; Pred. No. 5.8e+02;  
 Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 19 FXXXXXXGYXF 29  
 DB 98 FTASKNGTYKF 108  
 RESULT 7  
 Q9KGP4 PRELIMINARY; PRT; 250 AA.  
 AC Q9KGP4;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein BH0013.  
 GN BH0013.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 NCBI\_TaxID=6665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Maeni N.,  
 RA Fuji F., Hitama C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis."  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 DR EMBL; AP001507; BA003732.1; -;  
 DR InterPro: IPR006520; Tail\_N.

DR TIGRFA6; TIGR01633; phi3626\_gpi4\_N; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 250 AA; 28481 MW; DCD4B472571070FB CRC64;  
 Query Match 38.6%; Score 22; DB 16; Length 250;  
 Best Local Similarity 36.4%; Pred. No. 6.5e+02;  
 Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 19 FXXXXXXGYXF 29  
 DB 132 FTESGAGSYDF 142  
 RESULT 8  
 Q8CS05 PRELIMINARY; PRT; 290 AA.  
 AC Q8CS05;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Conserved hypothetical protein.  
 GN SEI006.  
 OS Staphylococcus epidermidis.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 NCBI\_TaxID=1282;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 12228;  
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,  
 RA Chen Z., Wen Y.;  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB016747; AA004603.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 290 AA; 33745 MW; A13B6A424943B195 CRC64;  
 Query Match 38.6%; Score 22; DB 16; Length 290;  
 Best Local Similarity 36.4%; Pred. No. 7.6e+02;  
 Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 19 FXXXXXXGYXF 29  
 DB 171 FKTTRSGVYTF 181  
 RESULT 9  
 Q9CNH6 PRELIMINARY; PRT; 296 AA.  
 AC Q9CNH6;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein PM0455.  
 GN PM0455.  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Pasteurella.  
 NCBI\_TaxID=747;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Pm70;  
 RX MEDLINE=21145866; PubMed=11248100;  
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.;  
 RT "Complete genomic sequence of Pasteurella multocida Pm70."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
 DR EMBL; AB006081; AA02539.1; -;  
 DR InterPro: IPR006314; DYP\_peroxidase.  
 DR TIGRFA6; TIGR01413; DYP\_perox\_fam; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 296 AA; 32951 MW; C4B616PAB4E40197 CRC64;  
 Query Match 38.6%; Score 22; DB 16; Length 296;  
 Best Local Similarity 36.4%; Pred. No. 7.8e+02;  
 Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 276 FTKAVTGSYYF 286

## RESULT 10

Q8DBA0 PRELIMINARY; PRT; 300 AA.  
AC Q8DBA0;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Tyra protein.  
GN V11922.  
OS Vibrio vulnificus.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrio.  
NCBI\_TaxID=672;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CMCP6;  
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
RA Choy H.E.;  
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";  
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE016803; AAC01033.1; -.  
KM Complete proteome.  
SQ SEQUENCE 300 AA; 33262 MW; 7369A026F23B6298 CRC64;

Query Match 38.6%; Score 22; DB 16; Length 300;  
Best Local Similarity 36.4%; Pred. No. 7.9e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 277 FTKAVTGAIFYF 287

## RESULT 11

Q9K059 PRELIMINARY; PRT; 302 AA.  
AC Q9K059;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Tyra protein.  
GN V02145.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrio.  
NCBI\_TaxID=666;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=El Tor N16961 / Serotype O1;  
RX MEDLINE=20406833; PubMed=10952301;  
RA Hedberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoti I., Sellers P.,  
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
RA Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
cholerae.";  
RL Nature 406:477-483(2000).  
DR EMBL; AE004287; AAF95290.1; -.  
DR TIGR; VC2145; -.  
DR InterPro; IPR006314; DYP\_peroxidase.  
DR TrEMBL; TIGR01413; DYP\_perox\_fam; 1.  
KM Complete proteome.  
SQ SEQUENCE 302 AA; 33496 MW; B45BAC93888F2F7E CRC64;

Query Match 38.6%; Score 22; DB 16; Length 302;  
Best Local Similarity 36.4%; Pred. No. 7.9e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 277 FTKAVTGAIFYF 287

## RESULT 12

Q8PSM8 PRELIMINARY; PRT; 323 AA.  
AC Q8PSM8;  
DT 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE Conserved protein.  
GN M1181.  
OS Methanosarcina mazei (Methanosarcina frisia).  
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
OC Methanosarcinaceae; Methanosarcina.  
NCBI\_TaxID=2209;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Goe1 / Goe1 / ATCC BAA-199 / DSM 3647 / OCM 88;  
RX MEDLINE=22120827; PubMed=1215824;  
RA Deppenmeier U., Johann A., Hartach T., Merkl R., Schmitz R.A.,  
RA Martinez-Arias R., Henne A., Wierer A., Baumer S., Jacobi C.,  
RA Brueggemann H., Lienhard T., Christmann A., Boemecke M., Steckel S.,  
RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,  
RA Fritz H.-U., Gottschalk G.;  
RT "The genome of Methanosarcina mazei: evidence for lateral gene  
transfer between Bacteria and Archaea.";  
RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).  
DR EMBL; AE013348; AAM30877.1; -.  
KM Complete proteome.  
SQ SEQUENCE 323 AA; 36826 MW; 8DDPE82C4AB87AE CRC64;

Query Match 38.6%; Score 22; DB 17; Length 323;  
Best Local Similarity 36.4%; Pred. No. 8.5e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 178 FAYVLSGAYF 188

## RESULT 13

Q8PSM8 PRELIMINARY; PRT; 325 AA.  
AC Q8PSM8;  
DT 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE Hypothetical protein MM3051.  
GN MM3051.  
OS Methanosarcina mazei (Methanosarcina frisia).  
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
OC Methanosarcinaceae; Methanosarcina.  
NCBI\_TaxID=2209;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Goe1 / Goe1 / ATCC BAA-199 / DSM 3647 / OCM 88;  
RX MEDLINE=22120827; PubMed=1215824;  
RA Deppenmeier U., Johann A., Hartach T., Merkl R., Schmitz R.A.,  
RA Martinez-Arias R., Henne A., Wierer A., Baumer S., Jacobi C.,  
RA Brueggemann H., Lienhard T., Christmann A., Boemecke M., Steckel S.,  
RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,  
RA Fritz H.-U., Gottschalk G.;  
RT "The genome of Methanosarcina mazei: evidence for lateral gene  
transfer between Bacteria and Archaea.";  
RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).  
DR EMBL; AE013360; AAM32747.1; -.



KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 325 AA; 36804 MW; 4BA35BD04B1D61CA CRC64;

Query Match  
 Best Local Similarity 38.6%; Score 22; DB 17; Length 325;  
 Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 19 FXXXXXXGYXF 29  
 Db 178 FAYVLSGAYAF 188

RESULT 14  
 Q966G5 PRELIMINARY; PRT; 355 AA.  
 AC Q966G5;  
 DT 01-DEC-2001 (T-EMBLrel. 19, Created)  
 DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
 DE Hypothetical protein R13D7.10.  
 GN R13D7.10.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None.  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Murray J., Wohlmann P., Elliott G.;  
 RT "The sequence of C. elegans cosmid R13D7.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RT "Direct Submission.";  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC006680; AAK72298.1;  
 DR WormBep; R13D7.10; CE18155.  
 DR InterPro; IPR002651; DUF32.  
 DR Pfam; PF01748; DUF32; 1.  
 SQ SEQUENCE 355 AA; 40110 MW; DF4CF926589473BF CRC64;

Query Match  
 Best Local Similarity 38.6%; Score 22; DB 5; Length 355;  
 Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 19 FXXXXXXGYXF 29  
 Db 251 FHATVAGLYSF 261

RESULT 15  
 Q8EJ24 PRELIMINARY; PRT; 384 AA.  
 AC Q8EJ24;  
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
 DE Outer membrane porin, putative.  
 GN S00312.  
 OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadaceae; Shewanella.  
 OX NCBI\_TaxID=70863;

RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NR-1;  
 RX MEDLINE=22297686; PubMed=12368813;  
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,  
 RA Weyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
 RA Madupu R., Peterson J.D., Mayhew L.A., White O., Wolf A.M.,  
 RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,  
 RA Mueller J., Khouri H., Gill J., Uppertack T.R., McDonald L.A.,  
 RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.,  
 RT Genome sequence of the dissimilatory metal ion-reducing bacterium  
 RT Shewanella oneidensis.";  
 RL Nat. Biotechnol. 20:1118-1123(2002).  
 DR EMBL; AE015479; AAN53397.1; -  
 DR TIGR; S00312; -  
 KW Porin; Complete proteome.  
 SQ SEQUENCE 384 AA; 42311 MW; 81B68F1A06C15B84 CRC64;

Query Match  
 Best Local Similarity 38.6%; Score 22; DB 16; Length 384;  
 Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 19 FXXXXXXGYXF 29  
 Db 144 FAGYSGAYAF 154

Search completed: January 12, 2004, 08:19:19  
 Job time : 9.26946 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:09 ; Search time 4.67166 Seconds  
(without alignments)  
1053.272 Million cell updates/sec

Title: US-10-085-167-3

Perfect score: 57  
Sequence: 1 FXXXXXXXXXXXXXXXXXXXXXGXFXFX 31

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19jun03.\*

```
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	38.6	163	19	AAW70458 Human T1-receptor
2	22	38.6	188	22	AAW93624 Human polypeptide,
3	22	38.6	215	19	AAW70459 Human T1-receptor
4	22	38.6	215	23	ABP65162 Hypoxia-regulated
5	22	38.6	219	22	ABBI11780 Human T1 receptor
6	22	38.6	224	22	AAW93844 Human polypeptide,
7	22	38.6	224	22	AAW92820 Human protein sequ
8	22	38.6	231	24	ABR41674 Human DITRP cell m
9	22	38.6	266	22	AAW82486 S. epidermidis ope

10	22	38.6	292	23	ABP39297 Staphylococcus epi
11	22	38.6	540	22	AAU57054 Propionibacterium
12	21	36.8	20	22	AAW73384 T. haxianum beta-
13	21	36.8	111	20	AAW89961 Antigen from clust
14	21	36.8	128	23	ABG76563 HCV E1 antigen mon
15	21	36.8	151	20	AAW37071 Protein which is s
16	21	36.8	154	24	ABP75824 Human secretory po
17	21	36.8	156	20	AAW35280 Chlamydia pneumoni
18	21	36.8	169	23	ABP27601 Streptococcus poly
19	21	36.8	183	23	ABP69421 Human polypeptide
20	21	36.8	201	22	AAW74750 Human secreted pro
21	21	36.8	201	22	AAW74759 Human secreted pro
22	21	36.8	201	22	AAW74760 Human secreted pro
23	21	36.8	201	23	ABG65324 Human albumin fusi
24	21	36.8	201	23	ABG65326 Human albumin fusi
25	21	36.8	201	23	ABG65327 Human albumin fusi
26	21	36.8	208	23	ABP69422 Human polypeptide
27	21	36.8	210	18	AAW20426 H. pylori surface
28	21	36.8	215	18	AAW20968 H. pylori transmem
29	21	36.8	254	22	AAU31151 Novel human secret
30	21	36.8	271	21	AAW07614 Arabidopsis thalia
31	21	36.8	271	21	AAW07613 Arabidopsis thalia
32	21	36.8	277	21	AAW07612 Arabidopsis thalia
33	21	36.8	277	21	AAW51317 Arabidopsis thalia
34	21	36.8	280	23	ABP29735 Streptococcus poly
35	21	36.8	281	22	ABG23769 Novel human diagno
36	21	36.8	295	21	AAW51316 Arabidopsis thalia
37	21	36.8	296	21	AAW07612 Arabidopsis thalia
38	21	36.8	300	21	AAW9317 Group B Streptococ
39	21	36.8	300	23	ABP26518 Streptococcus poly
40	21	36.8	307	21	AAW1832 Plasmodium falcipa
41	21	36.8	336	23	ABP62884 Human polypeptide
42	21	36.8	382	22	ABW60990 Drosophila melanog
43	21	36.8	408	22	ABW79473 Corynebacterium gl
44	21	36.8	441	22	AAW92180 C glutamicum prote
45	21	36.8	473	23	ABW98126 Human PPMV incyte

#### ALIGNMENTS

RESULT 1	AAW70458	AAW70458 standard; Protein; 163 AA.
ID	AAW70458	
XX	AAW70458;	
AC	AAW70458;	
DT	10-DEC-1998	(first entry)
XX		
DE	Human T1-receptor ligand III splice variant 1.	
XX		
KW	Human T1-receptor ligand III; T1-R ligand III; atherosclerosis;	
KW	autoimmune disease; inflammation; metabolic dysfunction;	
KW	immune-regulated disorder.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..24
FT	Protein	/note= "Signal peptide"
FT		25..163
FT		/note= "T1-R ligand III splice variant 1"
PN	MO9838311-A1.	
XX		
PD	03-SEP-1998.	
XX		
PF	26-FEB-1998;	98WO-US03483.
XX		
PR	28-FEB-1997;	97US-0039483.
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		

PI Gentz RL, Ni J, Ruben SM,  
XX WPI; 1998-495398/42.  
DR N-PSDB; AAV33461.  
XX  
PT New human T1-receptor ligand III splice variant(s) - useful for,  
PT e.g. detecting and creating immune system related disorders such as  
PT cancer and inflammation  
XX  
PS Claim 1; Fig 1; 115pp; English.  
XX  
CC The invention provides novel human T1-receptor ligand III (T1-R  
CC ligand III) splice variants and antibodies raised against these  
CC proteins. The present sequence represents the T1-R ligand III splice  
CC variant 1 protein sequence. The T1-R ligand III splice variants are  
CC claimed useful for screening agonists and antagonists. They are also  
CC claimed useful for treating disorders such as atherosclerosis,  
CC autoimmune disease, inflammation, metabolic dysfunction and  
CC immune-regulated disorders.  
XX  
SQ Sequence 163 AA;  
XX  
Query Match 38.6%; Score 22; DB 19; Length 163;  
Best Local Similarity 36.4%; Pred. No. 2.1e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 19 FXXXXXGXYXF 29  
DB 89 FTASKNGTYKF 99  
XX  
RESULT 2  
AAM93624  
ID AAM93624 standard; Protein; 188 AA.  
XX  
AC AAM93624;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide, SEQ ID NO: 3460.  
XX  
KW Human, full length cDNA; cDNA synthesis; oligo-capping.  
XX  
OS Homo sapiens.  
XX  
PN EP130094-A2.  
XX  
PD 05-SEP-2001.  
XX  
PF 07-JUL-2000; 2000EP-0114089.  
XX  
PR 08-JUL-1999; 99JP-0194486.  
PR 11-JAN-2000; 2000JP-0118774.  
PR 02-MAY-2000; 2000JP-0183765.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
PI Oca T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
XX WPI; 2001-524255/58.  
DR N-PSDB; AAK94557.  
XX  
PT 830 Primers useful for synthesizing full length cDNA clones and their  
PT use in genetic manipulation -  
XX  
PS Claim 8; SEQ ID NO 3460; 1380pp + sequence listing; English.  
XX  
CC The invention relates to primers for synthesizing full length cDNA  
CC clones. 830 cDNA molecules encoding a human protein have been  
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA  
CC molecules have been determined. Primers for synthesizing the full length  
CC cDNA are useful for clarifying the function of the protein encoded by

CC the cDNA. The full length clones were obtained by construction of full  
CC length enriched cDNA libraries that were synthesised by the oligo-capping  
CC method. The primers enable the production of the full length cDNA easily  
CC without any special methods. The present sequence is a polypeptide  
CC encoded by a full length human cDNA of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in CD-ROM format directly from EPO.  
XX  
SQ Sequence 188 AA;  
XX  
Query Match 38.6%; Score 22; DB 22; Length 188;  
Best Local Similarity 36.4%; Pred. No. 2.4e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 19 FXXXXXGXYXF 29  
DB 98 FTASKNGTYKF 108  
XX  
RESULT 3  
AAM70459  
ID AAM70459 standard; Protein; 215 AA.  
XX  
AC AAM70459;  
XX  
DT 10-DEC-1998 (first entry)  
XX  
DE Human T1-receptor ligand III splice variant 2.  
XX  
KW Human T1-receptor ligand III; T1-R ligand III; atherosclerosis;  
KW autoimmune disease; inflammation; metabolic dysfunction;  
KW immune-regulated disorder.  
XX  
OS Homo sapiens.  
XX  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..24  
FT /note= "Signal peptide"  
FT Protein 25..215  
FT /note= "T1-R ligand III splice variant 2"  
XX  
PN WO9838311-A1.  
XX  
PD 03-SEP-1998.  
XX  
PF 26-FEB-1998; 98WO-US03483.  
XX  
PR 28-FEB-1997; 97US-0039483.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Gentz RL, Ni J, Ruben SM;  
XX WPI; 1998-495398/42.  
DR N-PSDB; AAV33462.  
XX  
PT New human T1-receptor ligand III splice variant(s) - useful for,  
PT e.g. detecting and creating immune system related disorders such as  
PT cancer and inflammation  
XX  
PS Claim 1; Fig 2; 115pp; English.  
XX  
CC The invention provides novel human T1-receptor ligand III (T1-R  
CC ligand III) splice variants and antibodies raised against these  
CC proteins. The present sequence represents the T1-R ligand III splice  
CC variant 2 protein sequence. The T1-R ligand III splice variants are  
CC claimed useful for screening agonists and antagonists. They are also  
CC claimed useful for treating disorders such as atherosclerosis,  
CC autoimmune disease, inflammation, metabolic dysfunction and  
CC immune-regulated disorders.  
XX  
SQ Sequence 215 AA;



CC Depending on their biological activities, polypeptides and nucleotides of  
CC the invention are useful for preventing, treating or ameliorating medical  
CC conditions, e.g., by protein or gene therapy. Such conditions include  
CC cancers, hematopoietic disorders (e.g., myeloid or lymphoid cell  
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
CC arterial ischemia, bone disorders (e.g., osteoporosis), and abnormal  
CC vascular growth. Polypeptides involved with tissue regeneration and  
CC repair (or nucleic acids encoding them) may be used to promote wound  
CC healing (e.g., of burns, incisions and ulcers), while those with  
CC immunomodulatory activities may be used in the treatment of viral,  
CC bacterial and fungal infections in addition to immune disorders.  
CC Polypeptides with growth factor activity may be used in cell cultures to  
CC promote cell growth. For example, such polypeptides may be used to  
CC manipulate stem cells in culture to give rise to neuroepithelial cells  
CC that can be used to augment or replace cells damaged by illness,  
CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
CC may also be used in the diagnosis of the above conditions, and in drug  
CC screening techniques. The present sequence represents a novel human  
CC polypeptide of the invention.

SQ Sequence 219 AA;

Query Match 38.6%; Score 22; DB 22; Length 219;

Best Local Similarity 36.4%; Pred. No. 2.8e+02; Mismatches 7; Indels 0; Gaps 0;

Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGXXF 29

Db 93 FTASKNGTYKF 103

RESULT 6

AAM93844

ID AAM93844 standard; Protein; 224 AA.

XX AAM93844;

DT 06-NOV-2001 (first entry)

XX Human polypeptide, SEQ ID NO: 3923.

XX Human; full length cDNA; cDNA synthesis; oligo-capping.

XX Homo sapiens.

XX EP130094-A2.

XX 05-SEP-2001.

XX 07-JUL-2000; 2000EP-0114089.

XX 08-JUL-1999; 99JP-0194486.

XX 11-JAN-2000; 2000JP-0118774.

XX 02-MAY-2000; 2000JP-0183765.

XX (HELI-) HELIX RES INST.

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ieshi S, Kawai Y;

XX Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX WPI; 2001-524255/58.

XX N-PSDB; AAK94800.

XX 830 Primers useful for synthesizing full length cDNA clones and their

XX use in genetic manipulation -

XX Claim 8; SEQ ID NO 3923; 1380bp + sequence listing; English.

XX The invention relates to primers for synthesizing full length cDNA

XX clones. 830 cDNA molecules encoding a human protein have been

XX isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA

XX molecules have been determined. Primers for synthesizing the full length

CC cDNA are useful for clarifying the function of the protein encoded by  
CC the cDNA. The full length clones were obtained by construction of full  
CC length enriched cDNA libraries that were synthesised by the oligo-capping  
CC method. The primers enable the production of the full length cDNA easily  
CC without any special methods. The present sequence is a polypeptide  
CC encoded by a full length human cDNA of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in CD-ROM format directly from EPO.

SQ Sequence 224 AA;

Query Match 38.6%; Score 22; DB 22; Length 224;

Best Local Similarity 36.4%; Pred. No. 2.9e+02; Mismatches 7; Indels 0; Gaps 0;

Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGXXF 29

Db 98 FTASKNGTYKF 108

RESULT 7

AAB92820

ID AAB92820 standard; Protein; 224 AA.

XX AAB92820;

DT 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:11348.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ieshi S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Claim 8; SEQ ID 11348; 2537bp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

XX full-length cDNAs defined in the specification. Where a primer set

XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary

XX to the complementary strand of a polynucleotide which comprises one of

XX the 5602 nucleotide sequences defined in the specification, where the

XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX of an oligonucleotide comprising a sequence complementary to the

XX complementary strand of a polynucleotide which comprises a 5'-end

XX sequence and an oligonucleotide comprising a sequence complementary to a

XX polynucleotide which comprises a 3'-end sequence, where the

XX oligonucleotide comprises at least 15 nucleotides and the combination of

XX the 5'-end sequence/3'-end sequence is selected from those defined in

XX the specification. The primer sets can be used in antisense therapy and

in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 224 AA:

Query Match 38.6%; Score 22; DB 22; Length 224;  
Best Local Similarity 36.4%; Pred. No. 2.9e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 19 FXXXXXGXGXF 29  
DB 98 FTASKNGTYKF 108

RESULT 8

ID ABR41674 standard; Protein; 231 AA.

AC ABR41674;

DT 02-JUN-2003 (first entry)

DE Human D1THP cell membrane protein.

Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;  
cancer; cell proliferative disorder; autoimmune disorder;  
inflammatory disorder; infection; hormonal disorder; metabolic disorder;  
neurological disorder; gastrointestinal disorder; transport disorder;  
connective tissue disorder; drug screening; proteome analysis;  
gene therapy; antisense therapy; genotyping; transgenic animal; knock in;  
disease model; toxicological testing; transcript imaging;  
cell membrane protein.

OS Homo sapiens.

PN WO200297031-A2.

PD 05-DEC-2002.

PE 27-MAR-2002; 2002WO-US10056.

PR 28-MAR-2001; 2001US-279619P.

PR 29-MAR-2001; 2001US-280067P.

PR 29-MAR-2001; 2001US-280068P.

PR 16-MAY-2001; 2001US-291280P.

PR 17-MAY-2001; 2001US-291829P.

PR 17-MAY-2001; 2001US-291849P.

PR 19-JUN-2001; 2001US-299428P.

PR 20-JUN-2001; 2001US-299776P.

PR 20-JUN-2001; 2001US-300011P.

PA (INCY-) INCYTE GENOMICS INC.

PI Daffo A, Jones AI, Tran AB, Dahl CR, Gietzen D, Chinn J;

PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshy SR;

PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;

PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;

PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;

XX Claim 27; SEQ ID No 1209; 591pp; English.

PS The invention relates to novel human diagnostic and therapeutic  
XX polynucleotides designated dithp (ACC46080-ACC46749) and to their  
CC encoded proteins (D1THP: ABR41136-ABR41812). The invention also relates  
CC to polynucleotide sequences at least 90% identical to the dithp cDNA  
CC sequences of the invention; recombinant vectors, host cells and  
CC transgenic organisms comprising a dithp nucleic acid sequence; the  
CC recombinant production of D1THP proteins; antibodies specific for D1THP  
CC proteins; microarrays comprising dithp nucleic acid sequences; methods  
CC of detecting dithp nucleotide and protein sequences; methods of screening  
CC for compounds which specifically bind a D1THP protein; and methods of  
CC assessing the toxicity of test compounds using a dithp hybridisation  
CC probe. Dithp nucleic acid sequences and D1THP proteins may be used in the  
CC diagnosis of a wide variety of conditions including cancer and other cell  
CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,  
CC viral, fungal or parasitic infections; hormonal disorders; metabolic  
CC disorders; neurological disorders; gastrointestinal disorders; transport  
CC disorders; and connective tissue disorders. They may also be used to  
CC screen for modulators of protein activity or gene expression. D1THP  
CC proteins can additionally be used in analysis of the proteome of a tissue  
CC or cell type and to induce antibodies. The dithp nucleic acids are  
CC additionally useful in somatic or germline gene therapy of the disorders  
CC mentioned above, as a source of antisense sequences, as a source of  
CC probes and primers, in genotyping and identification of individuals, in  
CC the generation of transgenic animal models of human disease or knock in  
CC humanised animals, in toxicological testing, and in transcript imaging.  
CC The present sequence represents a D1THP protein which is a cell membrane  
CC protein.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 231 AA:

Query Match 38.6%; Score 22; DB 24; Length 231;  
Best Local Similarity 36.4%; Pred. No. 3e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 19 FXXXXXGXGXF 29  
DB 105 FTASKNGTYKF 115

RESULT 9

ID AAG82486 standard; Protein; 266 AA.

AC AAG82486;

DT 03-SEP-2001 (first entry)

DE S. epidermidis open reading frame protein sequence SEQ ID NO:2066.

XX Staphylococcus epidermidis SRI strain; infection; diagnosis;

KW vaccination; endocarditis.

OS Staphylococcus epidermidis.

PN WO200134809-A2.

PD 17-MAY-2001.

PE 09-NOV-2000; 2000WO-US30782.

PR 09-NOV-1999; 99US-0164258.

PA (GLAXO) GLAXO GROUP LTD.

PI Kimberly WF;

DR WPI; 2001-316495/33.

DR N-PSDB; AAH53336.  
 XX  
 PT Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,  
 PT useful for vaccinating against infections, e.g. endocarditis -  
 XX  
 PS Claim 18: Page 560; 2188bp; English.  
 XX  
 CC AAH53304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AAG81454 to AAG8120, from *Staphylococcus epidermidis*.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce host cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH5090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH5091 to  
 CC AAH5098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.  
 XX  
 SQ Sequence 266 AA;  
 XX  
 Query Match 38.6%; Score 22; DB 22; Length 266;  
 Best Local Similarity 36.4%; Pred. No. 3.4e+02;  
 Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 19 FXXXXXXGXXF 29  
 Db 147 FKTRSGVYTF 157  
 XX  
 RESULT 10  
 ABP39297  
 ID ABP39297 standard; Protein: 292 AA.  
 XX  
 AC ABP39297;  
 XX  
 DT 24-JUL-2002 (first entry)  
 XX  
 DE *Staphylococcus epidermidis* ORF amino acid sequence SEQ ID NO:4142.  
 XX  
 KW *Staphylococcus epidermidis*; open reading frame; ORF; bacterial infection;  
 KW antibacterial; gene therapy.  
 XX  
 OS *Staphylococcus epidermidis*.  
 XX  
 PN US6380370-B1.  
 XX  
 PD 30-APR-2002.  
 XX  
 PF 13-AUG-1998; 98US-0134001.  
 XX  
 PR 14-AUG-1997; 97US-055779P.  
 PR 08-NOV-1997; 97US-064964P.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 PI Doucette-Stamm LA, Bush D;  
 XX  
 DR WPI; 2002-38125/41.  
 DR N-PSDB; ABN91842.  
 XX  
 PT Novel isolated nucleic acid encoding a *Staphylococcus epidermidis*  
 PT polypeptide, useful for diagnosing and treating bacterial infections -  
 XX

PS Disclosure; SEQ ID 4142; 267bp; English.  
 XX  
 CC ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences  
 CC can also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life  
 CC cycle or inhibit S. epidermidis infection.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC USPTO web site.  
 XX  
 SQ Sequence 292 AA;  
 XX  
 Query Match 38.6%; Score 22; DB 23; Length 292;  
 Best Local Similarity 36.4%; Pred. No. 3.7e+02;  
 Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 19 FXXXXXXGXXF 29  
 Db 173 FKTRSGVYTF 183  
 XX  
 RESULT 11  
 AAU57054  
 ID AAU57054 standard; Protein: 540 AA.  
 XX  
 AC AAU57054;  
 XX  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE *Propionibacterium acnes* immunogenic protein #17950.  
 XX  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS *Propionibacterium acnes*.  
 XX  
 PN WO200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US12865.  
 XX  
 PR 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'walsonneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 DR WPI; 2001-616774/71.  
 DR N-PSDB; AAS59579.  
 XX  
 PT *Propionibacterium acnes* polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX  
 PS Example 1; SEQ ID No 18249; 1069bp; English.  
 XX  
 CC Sequences AAU39105-AAU68017 represent *Propionibacterium acnes* immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory



CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of *P. acnes* in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for *P. acnes* proteins. These antibodies can be used to  
CC downregulate expression and activity of *P. acnes* polypeptides and  
CC therefore treat *P. acnes* infections. The antibodies may also be used as  
CC diagnostic agents for determining *P. acnes* presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 540 AA;  
Query Match 38.6%; Score 22; DB 22; Length 540;  
Best Local Similarity 36.4%; Pred. No. 6.9e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 160 FSAGSAGDYLF 170

RESULT 12  
AAB73384  
ID AAB73384 standard; peptide; 20 AA.  
XX  
XX AAB73384;  
AC  
XX  
XX 05-JUN-2001 (first entry)  
DT  
XX  
XX  
DE T. harzianum beta-(1,6)-endoglucanase BGN16.1 N-terminal peptide.  
XX  
XX Beta-(1,6)-endoglucanase; BGN16.1; strain CECT 2413;  
KM antifungal; microbial cell wall degradation; antimicrobial;  
KW yeast extract production; disinfection; anti-biofilm;  
XX textile treatment; N-terminal peptide.  
XX  
XX Trichoderma harzianum.  
OS  
XX  
XX WO200109295-A1.  
PN  
XX  
XX 08-FEB-2001.  
PD  
XX  
XX 28-JUL-2000; 2000MO-ES00293.  
PF  
XX  
XX 31-JUL-1999; 99BS-0001747.  
PR  
XX  
XX (NEMB-) NEMBIOTECHNIC SA.  
PA (UYSE-) UNIV SEVILLA.  
PA (UYSA-) UNIV SALAMANCA.  
XX  
XX  
PI Montero Macario M, Rey Barrera M, Monte Vazquez E;  
PI Llobelo Gonzalez A;  
XX  
XX WPI; 2001-182948/18.  
DR  
XX  
XX  
PT New enzyme with beta-(1,6)-endoglucanase activity, useful e.g. as  
PT antifungal agent in animals or plants, from Trichoderma harzianum, also  
PT related nucleic acid -  
XX  
XX  
XX Example 13; Page 32; 62pp; Spanish.

CC The invention relates to novel Trichoderma harzianum CECT 2413 beta-  
CC (1,6)-endoglucanases, designated BGN16.1 and BGN16.3 (AAB73382 and  
CC AAB73383, respectively) and to nucleic acids encoding them (AAF76368 and  
CC AAF76269, respectively). The enzymes of the invention have antifungal  
CC activity, being able to degrade degrade beta-(1,6)-glucan components in  
CC microbial cell walls. The beta-(1,6)-endoglucanases are used as  
CC antifungal agents for the treatment and/or prevention of infections on  
CC plants, animals and humans, and to control species that contaminate foods

CC or harvested crops. They may also be used for disinfection of livestock  
CC installations and to prevent contamination of samples for analysis. Other  
CC applications of the beta-(1,6)-endoglucanases include the production of  
CC protoplasts and yeast extracts; extraction of mannoproteins; production  
CC of wine and fruit juices (improving filtration); eliminating dental  
CC plaque; cleaning teeth, dentures and contact lenses; removal of biofilms  
CC or fungi from coatings; and for the treatment of textiles, e.g., to  
CC remove excess dye. The present sequence represents a Trichoderma  
CC harzianum CECT 2413 beta-(1,6)-endoglucanase BGN16.1 N-terminal peptide  
CC used in the design of a degenerate PCR primer for the isolation of  
CC BGN16.1 cDNA.

SQ Sequence 20 AA;  
Query Match 36.8%; Score 21; DB 22; Length 20;  
Best Local Similarity 36.4%; Pred. No. 49;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 4 FASQDGRYQF 14

RESULT 13  
AAB89961  
ID AAB89961 standard; Protein; 111 AA.  
XX  
XX AAB89961;  
AC  
XX  
XX 20-MAR-2003 (updated)  
DT 18-FEB-1999 (first entry)  
DT  
XX  
XX Antigen from cluster 60.  
DE  
XX  
XX Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;  
KM peptic ulcer; gastric adenocarcinoma; gastric lymphoma.  
KW  
XX  
XX Helicobacter pylori.  
OS  
XX  
XX WO9849314-A2.  
PN  
XX  
XX 05-NOV-1998.  
PD  
XX  
XX 25-APR-1998; 98WO-US08487.  
PF  
XX  
XX 14-OCT-1997; 97US-0061958.  
PR 25-APR-1997; 97US-0045107.  
PR  
XX  
XX (GENE-) GENELABS TECHNOLOGIES INC.  
PA  
XX  
XX Chow TP, Fry KE, Lam MY, McAttee CP;  
PI WPI; 1999-009433/01.  
PI  
XX  
XX  
XX New Helicobacter pylori antigens and related nucleic acid sequences  
XX - useful in serological diagnosis and protective vaccines, providing  
XX long-lasting immune response  
XX  
XX  
XX Claim 1; Page 272; 402pp; English.

CC The present sequence represents a Helicobacter pylori antigenic protein  
CC that is characterised by immunoreactivity with H. pylori-positive  
CC antisera. The proteins are highly immunogenic and induce a long-lasting  
CC immune response that persists even after antimicrobial treatment. In  
CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are  
CC highly sensitive and specific. The specification also describes 69  
CC previously unrecognised immunogenic cluster families. H. pylori antigens  
CC are used to detect H. pylori-specific antibodies, for diagnosing  
CC infection or to confirm eradication of infection, and in vaccines to  
CC protect against H. pylori infection and related diseases (gastritis,  
CC peptic ulcer, gastric adenocarcinoma/lymphoma).  
CC (Updated on 20-MAR-2003 to correct PF field.)  
XX

SQ Sequence 111 AA;  
Query Match 36.8%; Score 21; DB 20; Length 111;  
Best Local Similarity 36.4%; Pred. No. 2,7e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 19 FXXXXXXGXXF 29  
DB 75 FGAVFTGFYAF 85  
RESULT 14  
ABG76563  
ID ABG76563 standard; Protein; 128 AA.  
XX  
AC ABG76563;  
XX  
XX 05-NOV-2002 (first entry)  
XX  
DE HCV E1 antigen monoclonal antibody #51.  
XX  
XX Human; HCV; hepatitis C virus; E1 antigen; monoclonal antibody; vaccine;  
KM hepatotropic; Fab; hypervariable region; E2 antigen; antibody.  
XX  
XX Homo sapiens.  
OS  
PN WO200260954-A1.  
XX  
PD 08-AUG-2002.  
XX  
PF 14-JAN-2002; 2002WO-SE00044.  
XX  
PR 12-JAN-2001; 2001US-260889P.  
XX  
PA (KARO-) KAROLINSKA INNOVATIONS AB.  
XX  
PI Drakenberg K, Persson MAA;  
XX  
DR WPI; 2002-608502/65.  
XX  
PT Vaccine comprising a human monoclonal antibody against hepatitis C  
PT virus (HCV) E1 or E2 antigen, useful for treating or preventing HCV  
PT infection -  
XX  
XX Disclosure; Page 55-56; 64pp; English.  
XX  
PS The invention relates to a human monoclonal antibody or its antigen  
CC binding fragments, which exhibit immunological binding affinity for a  
CC hepatitis C virus (HCV) E1 antigen, comprising an amino acid sequence  
CC homologous to the binding portion of a human antibody Fab molecule from a  
CC combinatorial antibody library. The vaccine composition comprising the  
CC antibodies or antigen binding fragments against HCV E1 or E2 antigen or  
CC its hypervariable region is useful in treating or preventing HCV  
CC infection in a subject. Sequences ABG76513-ABG76568 represent human  
CC monoclonal antibodies against HCV E1 antigen.  
XX  
SQ Sequence 128 AA;  
Query Match 36.8%; Score 21; DB 23; Length 128;  
Best Local Similarity 36.4%; Pred. No. 3.1e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 19 FXXXXXXGXXF 29  
DB 105 FVGGTTGYTF 115  
RESULT 15  
AAV37071  
ID AAV37071 standard; Protein; 151 AA.  
XX  
AC AAV37071v  
XX

DT 07-OCT-1999 (first entry)  
XX  
DE Protein which is specific to Chlamydia trachomatis.  
XX  
XX  
KM Vaccine; eye disease; conventional trachoma; nongonococcal urethritis; genital diseases; perithenitis;  
KM paratrachoma; inclusion conjunctivitis; genital diseases; perithenitis;  
KM nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
KM Bartholinitis; pneumopathy; venereal lymphogranulomatosis.  
XX  
XX Chlamydia trachomatis.  
XX  
PN WO928475-A2.  
XX  
PD 10-JUN-1999.  
XX  
XX  
PF 27-NOV-1998; 98WO-IB01939.  
XX  
PR 04-NOV-1998; 98US-0107077.  
PR 28-NOV-1997; 97FR-0015041.  
PR 17-DEC-1997; 97FR-0016034.  
XX  
XX (GENSET ) GENSET.  
PA  
XX  
PI Griffais R;  
XX  
XX WPI; 1999-371125/31.  
DR  
XX  
XX Genome sequence of Chlamydia trachomatis  
PT  
XX  
PS Disclosure; Page 883-884; 1755pp; English.  
XX  
XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome  
CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as  
CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
CC can also be used to control growth of the microorganism. Chlamydia  
CC trachomatis is responsible for a large number of diseases, e.g. eye  
CC diseases such as conventional trachoma, nongonococcal urethritis,  
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
CC perithenitis, Bartholinitis; pneumopathy in breast feeding infants;  
CC and venereal lymphogranulomatosis. The polypeptides of the invention  
CC may be of use in treating these diseases.  
XX  
SQ Sequence 151 AA;  
Query Match 36.8%; Score 21; DB 20; Length 151;  
Best Local Similarity 36.4%; Pred. No. 3.7e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 19 FXXXXXXGXXF 29  
DB 33 FLSESGKYL 43

Search completed: January 12, 2004, 08:15:40  
Job time : 6.67166 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:15:44 ; Search time 3.77445 Seconds  
(without alignments)  
1655.023 Million cell updates/sec

Title: US-10-085-167-3

Perfect score: 57 FXXXXXXXXXXXXXXXXXXXXXGXFFXX 31

Sequence: 1 FXXXXXXXXXXXXXXXXXXXXXGXFFXX 31

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747907 seqs, 201509753 residues

Total number of hits satisfying chosen parameters: 747907

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA:\*

1: /cgn2\_6/ptodaca/1/pubpaa/US07\_PUBCOMB.pep:\*

2: /cgn2\_6/ptodaca/1/pubpaa/PCT\_NEW\_PUB.pep:\*

3: /cgn2\_6/ptodaca/1/pubpaa/US06\_NEW\_PUB.pep:\*

4: /cgn2\_6/ptodaca/1/pubpaa/US06\_PUBCOMB.pep:\*

5: /cgn2\_6/ptodaca/1/pubpaa/US07\_NEW\_PUB.pep:\*

6: /cgn2\_6/ptodaca/1/pubpaa/PCTUS\_PUBCOMB.pep:\*

7: /cgn2\_6/ptodaca/1/pubpaa/US08\_NEW\_PUB.pep:\*

8: /cgn2\_6/ptodaca/1/pubpaa/US08\_PUBCOMB.pep:\*

9: /cgn2\_6/ptodaca/1/pubpaa/US09\_PUBCOMB.pep:\*

10: /cgn2\_6/ptodaca/1/pubpaa/US09C\_PUBCOMB.pep:\*

11: /cgn2\_6/ptodaca/1/pubpaa/US09\_NEW\_PUB.pep:\*

12: /cgn2\_6/ptodaca/1/pubpaa/US10\_PUBCOMB.pep:\*

13: /cgn2\_6/ptodaca/1/pubpaa/US10C\_PUBCOMB.pep:\*

14: /cgn2\_6/ptodaca/1/pubpaa/US10C\_PUBCOMB.pep:\*

15: /cgn2\_6/ptodaca/1/pubpaa/US10C\_PUBCOMB.pep:\*

16: /cgn2\_6/ptodaca/1/pubpaa/US10\_NEW\_PUB.pep:\*

17: /cgn2\_6/ptodaca/1/pubpaa/US60\_NEW\_PUB.pep:\*

18: /cgn2\_6/ptodaca/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	22	38.6	163	10	US-09-030-847-2
2	22	38.6	163	15	US-10-215-088-2
3	22	38.6	215	10	US-09-030-847-4
4	22	38.6	215	15	US-10-215-088-4
5	21	36.8	53	12	US-10-029-386-29069
6	21	36.8	201	14	US-10-039-865-6
7	21	36.8	201	14	US-10-039-865-7
8	21	36.8	201	15	US-10-060-255-59
9	21	36.8	201	15	US-10-060-255-68
10	21	36.8	201	15	US-10-060-255-69
11	21	36.8	300	12	US-09-769-736-86
12	21	36.8	420	12	US-10-369-493-11376
13	21	36.8	420	12	US-10-369-493-14778
14	21	36.8	420	12	US-10-369-493-14932
15	21	36.8	420	12	US-10-369-493-15269

16	21	36.8	421	12	US-10-369-493-13300	Sequence 12300, A
17	21	36.8	429	12	US-10-369-493-15506	Sequence 15506, A
18	21	36.8	429	12	US-10-369-493-15878	Sequence 15878, A
19	21	36.8	429	12	US-10-369-493-16255	Sequence 16255, A
20	21	36.8	441	10	US-09-738-626-5934	Sequence 5934, Ap
21	21	36.8	508	12	US-10-342-224-42	Sequence 42, Appl
22	21	36.8	661	12	US-10-369-493-10446	Sequence 10446, A
23	20	35.1	20	12	US-10-123-101-84	Sequence 84, Appl
24	20	35.1	20	12	US-10-310-734-114	Sequence 114, App
25	20	35.1	55	12	US-10-029-386-28388	Sequence 28388, A
26	20	35.1	98	12	US-10-023-634-48	Sequence 48, Appl
27	20	35.1	118	11	US-09-764-891-2701	Sequence 2701, Ap
28	20	35.1	142	11	US-09-880-505-47	Sequence 47, Appl
29	20	35.1	142	12	US-10-205-979-33	Sequence 33, Appl
30	20	35.1	142	14	US-10-051-643-47	Sequence 47, Appl
31	20	35.1	170	10	US-09-882-529-6	Sequence 6, Appl
32	20	35.1	170	10	US-09-882-529-8	Sequence 8, Appl
33	20	35.1	184	10	US-09-731-872-372	Sequence 372, App
34	20	35.1	184	12	US-09-876-997-372	Sequence 372, App
35	20	35.1	240	15	US-10-369-493-21423	Sequence 21423, A
36	20	35.1	241	15	US-10-156-761-13266	Sequence 13266, A
37	20	35.1	259	12	US-10-369-493-13403	Sequence 13403, A
38	20	35.1	292	10	US-09-738-626-6688	Sequence 6688, Ap
39	20	35.1	312	12	US-10-017-165-614	Sequence 614, App
40	20	35.1	312	12	US-10-292-798-540	Sequence 540, App
41	20	35.1	317	9	US-09-925-299-853	Sequence 853, App
42	20	35.1	317	11	US-09-925-299-853	Sequence 853, App
43	20	35.1	323	12	US-10-015-115-55	Sequence 55, Appl
44	20	35.1	323	12	US-10-369-493-5321	Sequence 5321, Ap
45	20	35.1	323	12	US-10-080-334-262	Sequence 262, App

## ALIGNMENTS

RESULT 1

US-09-030-847-2

Sequence 2, Application US/09030847A

Patent No. US2002011472A1

GENERAL INFORMATION:

APPLICANT: Nt, Ujan

TITLE OF INVENTION: T1-R Ligand III

FILE REFERENCE: PF357

CURRENT APPLICATION NUMBER: US/09/030,847A

CURRENT FILING DATE: 1998-02-26

EARLIER APPLICATION NUMBER: 60/039,483

EARLIER FILING DATE: 1997-02-28

NUMBER OF SEQ ID NOS: 16

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2

LENGTH: 163

TYPE: PRT

ORGANISM: Homo sapiens

US-09-030-847-2

Query Match 38.6%; Score 22; DB 10; Length 163;

Best Local Similarity 36.4%; Pred. No. 2.7e+02;

Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXXXX 29

Db 89 FTASKNGTYKF 99

RESULT 2

US-10-215-088-2

Sequence 2, Application US/10215088

Publication No. US20030069409A1

GENERAL INFORMATION:

APPLICANT: Nt et al.

TITLE OF INVENTION: T1-R Ligand III

FILE REFERENCE: PF357C1

CURRENT APPLICATION NUMBER: US/10/215,088

CURRENT FILING DATE: 2002-08-09  
PRIOR APPLICATION NUMBER: 09/030,847  
PRIOR FILING DATE: 1998-02-26  
PRIOR APPLICATION NUMBER: 60/039,483  
PRIOR FILING DATE: 1997-02-28  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 2  
LENGTH: 163  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-215-088-2

Query Match 38.6%; Score 22; DB 15; Length 163;  
Best Local Similarity 36.4%; Pred. No. 2.7e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 19 FXXXXXGXXF 29  
DB 89 FTASKNGTYKF 99

RESULT 3  
US-09-030-847-4  
Sequence 4, Application US/09030847A  
Patent No. US2002011472A1  
GENERAL INFORMATION:  
APPLICANT: NI, Jian  
TITLE OF INVENTION: T1-R Ligand III  
FILE REFERENCE: P357  
CURRENT APPLICATION NUMBER: US/09/030,847A  
CURRENT FILING DATE: 1998-02-26  
EARLIER APPLICATION NUMBER: 60/039,483  
EARLIER FILING DATE: 1997-02-28  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 4  
LENGTH: 215  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-030-847-4

Query Match 38.6%; Score 22; DB 10; Length 215;  
Best Local Similarity 36.4%; Pred. No. 3.5e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 19 FXXXXXGXXF 29  
DB 89 FTASKNGTYKF 99

RESULT 4  
US-10-215-088-4  
Sequence 4, Application US/10215088  
Publication No. US20030069409A1  
GENERAL INFORMATION:  
APPLICANT: NI et al.  
TITLE OF INVENTION: T1-R Ligand III  
FILE REFERENCE: P357C1  
CURRENT APPLICATION NUMBER: US/10/215,088  
CURRENT FILING DATE: 2002-08-09  
PRIOR APPLICATION NUMBER: 09/030,847  
PRIOR FILING DATE: 1998-02-26  
PRIOR APPLICATION NUMBER: 60/039,483  
PRIOR FILING DATE: 1997-02-28  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 4  
LENGTH: 215  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-215-088-4

Query Match 38.6%; Score 22; DB 15; Length 215;  
Best Local Similarity 36.4%; Pred. No. 3.5e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 19 FXXXXXGXXF 29  
DB 89 FTASKNGTYKF 99

RESULT 5  
US-10-029-386-29069  
Sequence 6, Application US/10029386  
Publication No. US2003019470A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GI  
FILE REFERENCE: AEOMICA-X-2  
CURRENT APPLICATION NUMBER: US/10/029,386  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 34288  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 29069  
LENGTH: 53  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO CHR7.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.75  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.8  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.65  
OTHER INFORMATION: SWISSPROT HIT: P34562, EVALU6 6.20e+00  
US-10-029-386-29069

Query Match 36.8%; Score 21; DB 12; Length 53;  
Best Local Similarity 36.4%; Pred. No. 1.7e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 19 FXXXXXGXXF 29  
DB 15 FNDSRKGSYGF 25

RESULT 6  
US-10-039-865-6  
Sequence 6, Application US/10039865  
Publication No. US20020132999A1  
GENERAL INFORMATION:  
APPLICANT: Goli, Surya K.  
TITLE OF INVENTION: NOVEL P24 VESICLE PROTEINS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/039,865  
FILING DATE: 19-Oct-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/215,736  
FILING DATE: <Unknown>  
APPLICATION NUMBER: 08/801,740  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0189 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX: <Unknown>  
SEQUENCE CHARACTERISTICS:  
LENGTH: 201 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: Genbank  
CLONE: 1212965  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-10-039-865-6

Query Match 36.8%; Score 21; DB 14; Length 201;  
Best Local Similarity 36.4%; Pred. No. 5.9e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 82 FAHMDGYKF 92

RESULT 7  
US-10-039-865-7  
Sequence 7, Application US/10039865  
Publication No. US20020132999A1  
GENERAL INFORMATION:  
APPLICANT: Goll, Surya K.  
Bandman, Olga  
TITLE OF INVENTION: NOVEL P24 VESICLE PROTEINS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/039,865  
FILING DATE: 19-Oct-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/215,736  
FILING DATE: <Unknown>  
APPLICATION NUMBER: 08/801,740  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0189 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX: <Unknown>  
SEQUENCE ID NO: 7;  
SEQUENCE CHARACTERISTICS:

LENGTH: 201 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: Genbank  
CLONE: 1213221  
SEQUENCE DESCRIPTION: SEQ ID NO: 7:  
US-10-039-865-7

Query Match 36.8%; Score 21; DB 15; Length 201;  
Best Local Similarity 36.4%; Pred. No. 5.9e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 82 FAHMDGYKF 92

RESULT 8  
US-10-060-255-59  
Sequence 59, Application US/10060255  
Publication No. US20030113840A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 25 Human secreted proteins  
FILE REFERENCE: P2042P1  
CURRENT APPLICATION NUMBER: US/10/060,255  
PRIOR FILING DATE: 2002-02-01  
PRIOR APPLICATION NUMBER: 09/781,417  
PRIOR FILING DATE: 2001-02-13  
PRIOR APPLICATION NUMBER: PCT/US00/22325  
PRIOR FILING DATE: 2000-08-16  
PRIOR APPLICATION NUMBER: 60/149,182  
PRIOR FILING DATE: 1999-08-17  
NUMBER OF SEQ ID NOS: 86  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 59  
LENGTH: 201  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-060-255-59

Query Match 36.8%; Score 21; DB 15; Length 201;  
Best Local Similarity 36.4%; Pred. No. 5.9e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 82 FAHMDGYKF 92

RESULT 9  
US-10-060-255-68  
Sequence 68, Application US/10060255  
Publication No. US20030113840A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 25 Human secreted proteins  
FILE REFERENCE: P2042P1  
CURRENT APPLICATION NUMBER: US/10/060,255  
PRIOR FILING DATE: 2002-02-01  
PRIOR APPLICATION NUMBER: 09/781,417  
PRIOR FILING DATE: 2001-02-13  
PRIOR APPLICATION NUMBER: PCT/US00/22325  
PRIOR FILING DATE: 2000-08-16  
PRIOR APPLICATION NUMBER: 60/149,182  
PRIOR FILING DATE: 1999-08-17  
NUMBER OF SEQ ID NOS: 86  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 68  
LENGTH: 201  
TYPE: PRT

ORGANISM: Homo sapiens  
US-10-060-255-68

Query Match 36.8%; Score 21; DB 15; Length 201;  
Best Local Similarity 36.4%; Pred. No. 5.9e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 82 FAAMMDGYKF 92

RESULT 10  
US-10-060-255-69  
Sequence 69, Application US/10060255  
Publication No. US2003011840A1  
GENERAL INFORMATION:

APPLICANT: Rosen et al.  
TITLE OF INVENTION: 25 Human secreted proteins  
FILE REFERENCE: P2042P1  
CURRENT APPLICATION NUMBER: US/10/060,255  
CURRENT FILING DATE: 2002-02-01  
PRIOR APPLICATION NUMBER: 09/781,417  
PRIOR FILING DATE: 2001-02-13  
PRIOR APPLICATION NUMBER: PCT/US00/22325  
PRIOR FILING DATE: 2000-08-16  
PRIOR APPLICATION NUMBER: 60/149,182  
PRIOR FILING DATE: 1999-08-17  
NUMBER OF SEQ ID NOS: 86  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 69  
LENGTH: 201  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-060-255-69

Query Match 36.8%; Score 21; DB 15; Length 201;  
Best Local Similarity 36.4%; Pred. No. 5.9e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 82 FAAMMDGYKF 92

RESULT 11  
US-09-769-736-86  
Sequence 86, Application US/09769736  
Publication No. US2003013875A1  
GENERAL INFORMATION:  
APPLICANT: Microbial Technics Limited  
APPLICANT: Le Page, Richard W  
APPLICANT: Wells, Jeremy M  
APPLICANT: Hanniffy, Sean B  
TITLE OF INVENTION: Proteins  
FILE REFERENCE: PWC/P21089wo  
CURRENT APPLICATION NUMBER: US/09/769,736  
CURRENT FILING DATE: 2003-02-14  
PRIOR APPLICATION NUMBER: GB 9816335.5  
PRIOR FILING DATE: 1998-07-27  
PRIOR APPLICATION NUMBER: US 60/125163  
PRIOR FILING DATE: 1999-03-19  
NUMBER OF SEQ ID NOS: 212  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 86  
LENGTH: 300  
TYPE: PRT  
ORGANISM: Streptococcus agalactiae  
US-09-769-736-86

Query Match 36.8%; Score 21; DB 12; Length 300;  
Best Local Similarity 36.4%; Pred. No. 8.6e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 242 FASVETGYKF 252

RESULT 12  
US-10-369-493-11376  
Sequence 11376, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:

APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 11376  
LENGTH: 420  
TYPE: PRT  
ORGANISM: Agrobacterium tumefaciens  
US-10-369-493-11376

Query Match 36.8%; Score 21; DB 12; Length 420;  
Best Local Similarity 36.4%; Pred. No. 1.2e+03;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 201 FALASMGIVF 211

RESULT 13  
US-10-369-493-14778  
Sequence 14778, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 14778  
LENGTH: 420  
TYPE: PRT  
ORGANISM: Agrobacterium tumefaciens  
US-10-369-493-14778

Query Match 36.8%; Score 21; DB 12; Length 420;  
Best Local Similarity 36.4%; Pred. No. 1.2e+03;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 201 FALASMGIVF 211

RESULT 14

```
US-10-369-493-14932
; Sequence 14932, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 14932
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-14932

Query Match          36.8%; Score 21; DB 12; Length 420;
Best Local Similarity 36.4%; Pred. No. 1.2e+03;
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXGXKX 29
Db 201 FALASMGIVVF 211

RESULT 15
US-10-369-493-15269
; Sequence 15269, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15269
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-15269

Query Match          36.8%; Score 21; DB 12; Length 420;
Best Local Similarity 36.4%; Pred. No. 1.2e+03;
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXGXKX 29
Db 201 FALASMGIVVF 211
```

Search completed: January 12, 2004, 08:33:42  
Job time : 5.77445 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2004, 08:12:14 ; Search time 1.57784 Seconds  
(without alignments)  
831.284 Million cell updates/sec

Title: US-10-085-167-3  
Perfect score: 57  
Sequence: 1 FXXXXXXXXXXXXXXXXXXXXXGXFX 31

Scoring table: BLOSUM62  
Gapop 10.0 , Gapept 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PCUTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	38.6	268	4 US-09-252-991A-18273	Sequence 18273, A
2	22	38.6	292	4 US-09-134-001C-4142	Sequence 4142, A
3	22	38.6	305	4 US-09-328-352-6069	Sequence 6069, A
4	22	38.6	404	4 US-09-328-352-8182	Sequence 8182, A
5	21	36.8	156	4 US-09-198-452A-698	Sequence 698, A
6	21	36.8	201	2 US-08-801-740-6	Sequence 6, A
7	21	36.8	201	2 US-08-801-740-7	Sequence 7, A
8	21	36.8	201	3 US-08-801-740-6	Sequence 6, A
9	21	36.8	201	3 US-08-801-740-7	Sequence 7, A
10	21	36.8	252	4 US-09-252-991A-30813	Sequence 30813, A
11	21	36.8	1446	4 US-09-071-035-404	Sequence 404, A
12	21	36.8	1448	4 US-09-071-035-402	Sequence 402, A
13	20	35.1	63	2 US-08-637-759B-501	Sequence 501, A
14	20	35.1	63	2 US-08-637-759B-501	Sequence 501, A
15	20	35.1	63	4 US-09-201-945-501	Sequence 501, A
16	20	35.1	142	2 US-08-997-080-47	Sequence 47, A
17	20	35.1	142	2 US-08-997-362-47	Sequence 47, A
18	20	35.1	142	3 US-08-873-970-47	Sequence 47, A
19	20	35.1	142	3 US-09-095-855-47	Sequence 47, A
20	20	35.1	142	3 US-08-705-247A-47	Sequence 47, A
21	20	35.1	142	4 US-09-324-542-47	Sequence 47, A
22	20	35.1	142	4 US-09-205-426-47	Sequence 47, A
23	20	35.1	142	4 US-09-200-643-47	Sequence 47, A
24	20	35.1	143	4 US-08-685-808-2	Sequence 2, A
25	20	35.1	143	3 US-08-505-860C-2	Sequence 2, A
26	20	35.1	162	2 US-08-606-143-44	Sequence 44, A
27	20	35.1	336	4 US-08-987-691A-5	Sequence 5, A

28	20	35.1	408	4 US-09-252-991A-33131	Sequence 33131, A
29	20	35.1	469	4 US-09-198-452A-793	Sequence 793, A
30	20	35.1	479	4 US-09-996-243-216	Sequence 216, A
31	20	35.1	492	4 US-09-252-991A-27853	Sequence 27853, A
32	20	35.1	498	3 US-09-332-468A-18	Sequence 18, A
33	20	35.1	498	3 US-09-232-468A-24	Sequence 24, A
34	20	35.1	498	3 US-08-686-968C-231	Sequence 231, A
35	20	35.1	498	4 US-09-784-984B-52	Sequence 52, A
36	20	35.1	498	4 US-09-784-984B-54	Sequence 54, A
37	20	35.1	546	2 US-09-067-351-1	Sequence 1, A
38	20	35.1	546	3 US-09-360-490-1	Sequence 1, A
39	20	35.1	646	4 US-09-336-643A-10	Sequence 10, A
40	20	35.1	922	4 US-09-198-452A-15	Sequence 15, A
41	20	35.1	1430	3 US-09-008-172-2	Sequence 2, A
42	20	35.1	1430	3 US-09-210-361-6	Sequence 6, A
43	20	35.1	1430	4 US-09-740-274-6	Sequence 6, A
44	20	35.1	3177	2 US-08-477-451-4	Sequence 4, A
45	19	33.3	32	4 US-08-525-539A-35	Sequence 35, A

## ALIGNMENTS

RESULT 1  
US-09-252-991A-18273  
Sequence 18273, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 18273  
LENGTH: 268  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-18273

Query Match 38.6% Score 22; DB 4; Length 268;  
Best Local Similarity 36.4%; Pred. No. 1e+02; 7; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches

QY 19 FXXXXXXXXX 29  
DB 41 FLAVATGVAF 51

RESULT 2  
US-09-134-001C-4142  
Sequence 4142, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 4142  
LENGTH: 292  
TYPE: PRT

ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4142

Query Match 38.6%; Score 22; DB 4; Length 292;  
Best Local Similarity 36.4%; Pred. No. 1.1e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 173 FKTRSGVYTF 183

RESULT 3  
US-09-328-352-6069  
Sequence 6069, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 6069  
LENGTH: 305  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-6069

Query Match 38.6%; Score 22; DB 4; Length 305;  
Best Local Similarity 36.4%; Pred. No. 1.1e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 52 FTGTAGTYTF 62

RESULT 4  
US-09-328-352-8182  
Sequence 8182, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 8182  
LENGTH: 404  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-8182

Query Match 38.6%; Score 22; DB 4; Length 404;  
Best Local Similarity 36.4%; Pred. No. 1.5e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 242 FLKTTAGHYAF 252

RESULT 5  
US-09-198-452A-698  
Sequence 698, Application US/09198452A  
Patent No. 6559294  
GENERAL INFORMATION:  
APPLICANT: Griffiths, R.  
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments

TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
FILE OF INVENTION: and treatment of infection  
FILE REFERENCE: 9710-003-999  
CURRENT APPLICATION NUMBER: US/09/198,452A  
CURRENT FILING DATE: 1998-11-24  
NUMBER OF SEQ ID NOS: 6849  
SEQ ID NO 698  
LENGTH: 156  
TYPE: PRT  
ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-698

Query Match 36.8%; Score 21; DB 4; Length 156;  
Best Local Similarity 36.4%; Pred. No. 1.1e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29  
DB 88 FSLPSTGYTF 98

RESULT 6  
US-08-801-740-6  
Sequence 6, Application US/08801740  
Patent No. 5869639  
GENERAL INFORMATION:  
APPLICANT: Goli, Surya K.  
TITLE OF INVENTION: NOVEL P24 VESICLE PROTEINS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTED for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/801,740  
FILING DATE: Herewith  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0189 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:

INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 201 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: Genbank  
CLONE: 1212965  
US-08-801-740-6

Query Match 36.8%; Score 21; DB 2; Length 201;  
Best Local Similarity 36.4%; Pred. No. 1.4e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXXGYXF 29

Db 82 FAAHMDGTYKF 92

## RESULT 7

US-08-801-740-7

; Sequence 7, Application US/08801740

; Patent No. 5669639

; GENERAL INFORMATION:

; APPLICANT: Goli, Surya K.

; APPLICANT: Bandman, Olga

; TITLE OF INVENTION: NOVEL P24 VESICLE PROTEINS

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/801,740

; FILING DATE: Herewith

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0189 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 201 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: GenBank

; CLONE: 1213221

; US-08-801-740-7

; Query Match

; Best Local Similarity 36.8%; Score 21; DB 2; Length 201;

; Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

; Oy 19 FXXXXXGXYP 29

; Db 82 FAAHMDGTYKF 92

## RESULT 8

US-08-801-740-6

; Sequence 6, Application US/08801740

; Patent No. 6130325

; GENERAL INFORMATION:

; APPLICANT: Goli, Surya K.

; APPLICANT: Bandman, Olga

; TITLE OF INVENTION: NOVEL P24 VESICLE PROTEINS

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/801,740

FILING DATE: 14-Feb-1997

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: &lt;Unknown&gt;

FILING DATE: &lt;Unknown&gt;

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0189 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX: &lt;Unknown&gt;

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 201 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: 1212965

US-08-801-740-6

; Query Match

; Best Local Similarity 36.8%; Score 21; DB 3; Length 201;

; Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

; Oy 19 FXXXXXGXYP 29

; Db 82 FAAHMDGTYKF 92

## RESULT 9

US-08-801-740-7

; Sequence 7, Application US/08801740

; Patent No. 6130325

; GENERAL INFORMATION:

; APPLICANT: Goli, Surya K.

; APPLICANT: Bandman, Olga

; TITLE OF INVENTION: NOVEL P24 VESICLE PROTEINS

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/801,740

; FILING DATE: 14-Feb-1997

; CLASSIFICATION: &lt;Unknown&gt;

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: &lt;Unknown&gt;

; FILING DATE: &lt;Unknown&gt;

; ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0189 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 201 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1213221  
US-08-801-740-7  
SEQUENCE DESCRIPTION: SEQ ID NO: 7:  
  
Query Match 36.8%; Score 21; DB 3; Length 201;  
Best Local Similarity 36.4%; Pred. No. 1.4e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
QY 19 FXXXXXXGXXF 29  
DB 82 FAHMDGTXYKF 92  
  
RESULT 10  
US-09-252-991A-30813  
Sequence 30813, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 30813  
LENGTH: 252  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30813  
  
Query Match 36.8%; Score 21; DB 4; Length 252;  
Best Local Similarity 36.4%; Pred. No. 1.8e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
QY 19 FXXXXXXGXXF 29  
DB 167 FASEDSGVYRF 177  
  
RESULT 11  
US-09-071-035-404  
Sequence 404, Application US/09071035  
Patent No. 6448043  
GENERAL INFORMATION:  
APPLICANT: Gil H. Choi  
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 496  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA

ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/071,035  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: A. Anders Brookes  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB369P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8512  
TELEFAX: (301) 309-8504  
INFORMATION FOR SEQ ID NO: 404:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1416 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-071-035-404  
  
Query Match 36.8%; Score 21; DB 4; Length 1416;  
Best Local Similarity 36.4%; Pred. No. 1e+03;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
QY 19 FXXXXXXGXXF 29  
DB 959 FTTDEKQYSF 969  
  
RESULT 12  
US-09-071-035-402  
Sequence 402, Application US/09071035  
Patent No. 6448043  
GENERAL INFORMATION:  
APPLICANT: Gil H. Choi  
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 496  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/071,035  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: A. Anders Brookes  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB369P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8512  
TELEFAX: (301) 309-8504  
INFORMATION FOR SEQ ID NO: 402:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1448 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-071-035-402

Query Match 36.4%; Score 21; DB 4; Length 1448;  
Best Local Similarity 36.4%; Pred. No. 1.1e+03;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXGXYF 29  
DB 986 FTTDEKQYSF 996

RESULT 13  
US-08-637-759B-501  
Sequence 501, Application US/08637759B  
Patent No. 5876931  
GENERAL INFORMATION:  
APPLICANT: David William Holden  
TITLE OF INVENTION: Identification of Genes  
NUMBER OF SEQUENCES: 501  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
City: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/637,759B  
FILING DATE: 03-MAY-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB95/02875  
FILING DATE: 11-DEC-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: RPLMS 101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 873-8794  
TELEFAX: (404) 873-8795  
INFORMATION FOR SEQ ID NO: 501:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 63 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-08-637-759B-501

Query Match 35.1%; Score 20; DB 2; Length 63;  
Best Local Similarity 36.4%; Pred. No. 85;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXGXYF 29  
DB 4 FOSIVAGYYF 14

RESULT 14

US-08-871-355A-501  
Sequence 501, Application US/08871355A  
Patent No. 6015669  
GENERAL INFORMATION:  
APPLICANT: David William Holden  
TITLE OF INVENTION: Identification of Genes  
NUMBER OF SEQUENCES: 501  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
City: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/871,355A  
FILING DATE: 09-JUN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB95/02875  
FILING DATE: 11-DEC-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: RPLMS 101 CON  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 873-8794  
TELEFAX: (404) 873-8795  
INFORMATION FOR SEQ ID NO: 501:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 63 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-08-871-355A-501

Query Match 35.1%; Score 20; DB 3; Length 63;  
Best Local Similarity 36.4%; Pred. No. 85;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 FXXXXXGXYF 29  
DB 4 FOSIVAGYYF 14

RESULT 15  
US-09-201-945-501  
Sequence 501, Application US/09201945  
Patent No. 6142215  
GENERAL INFORMATION:  
APPLICANT: David William Holden  
TITLE OF INVENTION: Identification of Genes  
NUMBER OF SEQUENCES: 501  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
City: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/201,945
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/637,759
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabat, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 501:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 63 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; US-09-201-945-501

```

```

Query Match 35.1%; Score 20; DB 4; Length 63;
Best Local Similarity 36.4%; Pred. No. 85;
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

OY 19 FXXXXXKXYP 29
Db 4 FQIVAGYYF 14

```

Search completed: January 12, 2004, 08:20:09  
 Job time : 3.57784 secs